

```
*      124231 124330: gap of      100 bp
*      124331 159550: contig of 35220 bp in length.
FEATURES
    source
        1..159550
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /map="11"
            /clone="RP11-278E23"
            /clone_id="RPC1-11 Human Male BAC"
            1..1729
                misc_feature
                    /note="assembly_fragment"
                    1830..3716
                        misc_feature
                            /note="assembly_fragment"
                            3817..6950
                                misc_feature
                                    /note="assembly_fragment"
                                    7051..9136
                                        misc_feature
                                            /note="assembly_fragment"
                                            9237..17859
                                                misc_feature
                                                    /note="assembly_fragment"
                                                    17960..26939
                                                        misc_feature
                                                            /note="assembly_fragment"
                                                            27040..36269
                                                                misc_feature
                                                                    /note="assembly_fragment"
                                                                    36370..48574
                                                                        misc_feature
                                                                            /note="assembly_fragment"
                                                                            48675..62315
                                                                                misc_feature
                                                                                    /note="assembly_fragment"
                                                                                    62416..77457
                                                                                        misc_feature
                                                                                            /note="assembly_fragment
                                                                                            vector_end:SP6
                                                                                            vector_side:right"
                                                                                            77558..96884
                                                                                                misc_feature
                                                                                                    /note="assembly_fragment"
                                                                                                    96985..124230
                                                                                                        misc_feature
                                                                                                            /note="assembly_fragment"
                                                                                                            124331..159550
                                                                                                                misc_feature
                                                                                                                    /note="assembly_fragment
                                                                                                                    clone_end:T7
                                                                                                                    vector_side:right"
BASE COUNT      42186 a 37364 c 37475 g 41307 t 1218 others
ORIGIN
Query Match      70.0%: Score 349.8; DB 2; Length 159550;
Best Local Similarity 85.1%: Pred. No. 7.6e-24;
Matches 416; Conservative 0; Mismatches 67; Indels 6; Gaps 2;
QY      18 GAGGAGCGTGGCCAGGCGCATCGCTAATGAGGAGCGCCGACCAAGGGCATCGCTAATGAGGAC 77
      124879 GAGGAGCGCCGCCAGCGCATCGCCAGAGGAGACGCCGCCACGGCATCGCCAGCGAGGAC 124820
QY      78 ACCACCCAGTCGATCGCCAGCAGGAAAGCCGCCAGGGCGCATGC--CGAGAGCGCCATC 134
      124819 GCCGCGCAGGCGCATCGCCAGCAGGAGGAGCGCCGCCAGGGCATCGCCAGCGAGGAC 124760
QY      135 CAGGCGATCGCCAGCAGGAGGAGGTGCCCCAGGCGCATCGCCCAATGGGGTGGCGGACAGCGGC 194
      124759 CAGGCGATCGCCAGAGGAGGAGCGCCGCCAGGGCATCGCCCAACGAGGAGCGCCGCCAGGCG 124700
QY      195 ATCGCAATGAGAGCGCCAGCCAGGCGATCGCCCAACTGGAGCGCCGTCACGCGCTTCGCGC 254
      124699 ATCGCCACAGAGGCGCGCGCCAGGCGATCGCCCAAGAGGAGCGCGCCAGGGCATCGGC 124640
QY      255 AACGGGAGCGCCGTCCTCAGCTTGGCCAAAGGGAGCGCCGCCAGGGCATCGCCCAACGAG 314
      124639 AAGAGAGGCGCGCCCGCAGGCGATCGCCCAACGAGGAGCGCGCCAGGGCATCGCCCAACGAG 124580
QY      315 GAGCGCACCAAGGGCATGGGCAACGAGGTCCACCATCCAGGGCATGCTAAGAGGAGCGCC 374
      124579 GACTCCGCCACGGAATCGCCAGGAGGAGCGCCGCCACGGCATCGCCAGCGAGGAGCGCC 124520
QY      375 GTCCAGGCGCATCGCTAAGAGGTGGCGCCAGGCGCATCGCCCAACGAGGAGCGCCCGCAG 434
      124519 GTCCAGGCGCATCGCCCAAGGAGGAGCGCCGCCAGGGCATCGCCCAACGAGGAGCGCCCGCAG 124460
QY      435 GGAATCGCC--GAGATGTCGCACAGAGGGCATCGCCCAACGAGGAGCGCCCGCCAGGCATC 491
      124459 GGCATCGCCCAAGGAGGAGCGCCGCCAGGCGATCGCCCAAGGAGGAGCGCGCCAGGCGCATC 124400
QY      492 GCCAACAG 500
      124399 GCCAACGAG 124391
```

```
RESULT 11
AC008103/c      98557 bp      DNA      linear      PRI 27-MAY-2000
LOCUS
DEFINITION      Homo sapiens chromosome 22q12 clone pac699j1, complete sequence.
ACCESSION      AC008103
VERSION      AC008103.27      GI:7229751
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Homo sapiens Chromosome 22q12 PAC Clone pac699j1 in CES-DGCR Region
unpublished
2 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (22-JUL-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (22-JUL-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (08-OCT-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (08-DEC-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (17-DEC-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (26-JAN-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
8 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (01-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
9 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (03-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
```

REFERENCE
AUTHORS
TITLE
JOURNAL
OK 73019, USA
10 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (05-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
11 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (08-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
12 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (10-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
13 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (14-FEB-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
14 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (12-MAR-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
15 (bases 1 to 98557)
Wang,Q., Fang,F., Fu,Y., Pan,H., Edelman,L. and Roe,B.A.
Direct Submission
Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Mar 12, 2000 this sequence version replaced gi:6958024.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC008132(pac9506)136107 145670 (0) overlaps AC008103(pac699j1) 1
9536 (89001) AC007981(b291) 132871 175358 (0) overlaps
AC008103(pac699j1) 1 44355 (54202) AC007325(b48) 71505 165050 (0)
overlaps AC008103(pac699j1) 1 94144 (4413) AC008103(pac699j1) 56194
98557 (0) overlaps AC007326(p423) 1 42365 (60533).
FEATURES
source
1.98557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q12"
/clone="pac699j1"
BASE COUNT 25600 a 24617 c 24180 g 24160 t
ORIGIN
Query Match 69.6%; Score 348.2; DB 9; Length 98557;
Best Local Similarity 84.9%; Pred. No. 1.2e-23;
Matches 415; Conservative 0; Mismatches 68; Indels 6; Gaps 2;
QY 18 GAGGAGCGTCCCGAGGCGATCGCTAATGAGGAGCGCCAGCGGCGATCGCTAATGAGGAC 77
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27916 GAGGAGCGCGCCGACGAATCGCGAGGAGAGCGCCCGCCAGCGCATCGCCAGCGAGGAC 27857
QY 78 ACCACCGATGATCGCCAGCAAGAGGAGCGCCCGCCAGGCGATCGCC---GAGGAGCGCATC 134
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27856 GCCGCCAGGCGATCGCCAGCGAGGAGCGCGCCCGCCAGGCGATCGCCAGGAGAGAGCGCCGC 27797
QY 135 CAGGCGATCGCCAGGAGGAGGAGGTTGCCAGGCGATCGCCCATGGGGTCCGCCAGACAGGCG 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27796 CAGGCGATCGCCAGGAGGAGCGCGCCCGCCAGGCGATCGCCCAAGAGGAGCGCGCCAGGCG 27737
QY 195 ATCGCATGAGGAGCGCCAGGCGATCGCCACTGGGAGCGCGCGCTTCGAGCGCTTCGGC 254
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 27736 ATCGCCAAAGGAGAGCGCGCCGACGCGATCGCCAAAGAGAGACTCCGCGCCAGGAATCGCC 27677
QY 255 AAGGGGAGCCCGTCCATCGTTGCCAGCGGAGCGCGCCCGCCAGGCGATCGCCAAAGGCG 314
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27676 AGCGAGACGCCCGCCAGCGATCGCCAGGAGCGCGCTCGACGGCGATCGCCAAAGGAG 27617
QY 315 GAGCGCCAAAGGCGATCGGCGAACGAGTCCATCCACGCGATCGCTAACGAGAGCGCG 374
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27616 GAGCGCCCGCCAGGCGATCGCCAGGAGAGCGCGCCCGCCAGGCGATCGCCAAAGGAGAGCGCG 27557
QY 375 GTCCAGGCGATCGCTAACGAGTGGCGCCCGCCAGGCGATCGCCAGGAGAGCGCGCGCG 434
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27556 GCCAGGCGATCGCCAGGAGGAGCGCGCCCGCCAGGCGATCGCCAAAGGAGCGCGCGCGCG 27497
QY 435 GGAATGCGC---GAGGATGCGCACAGGCGCATCGCCCAAGAGAGAGCGCGCCAGGCGATC 491
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27496 GGCATCGCCAAAGAGAGAGCGCGCCCGCCAGGCGATCGCCAAAGAGAGAGCGCGCCAGGCGATC 27437
QY 492 GCCACAG 500
Db 27436 GCCACAG 27428
RESULT 12
AC007708/c 106650 bp DNA linear PRI 31-MAY-2000
LOCUS
DEFINITION
AC007708 Homo sapiens chromosome 22q11 clone b563b9, complete sequence.
ACCESSION
VERSION
KEYWORDS
AC007708.13 GI:5923689
HTG.
SOURCE
ORGANISM
human.
REFERENCE
AUTHORS
TITLE
JOURNAL
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 106650)
Zhan,M. and Roe,B.A.
Homo sapiens Chromosome 22q11 BAC Clone b563b9 In BCRU2-GGT Region
Unpublished
2 (bases 1 to 106650)
Zhan,M. and Roe,B.A.
Direct Submission
Submitted (02-JUN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 106650)
Zhan,M. and Roe,B.A.
Direct Submission
Submitted (25-SEP-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 106650)
Zhan,M. and Roe,B.A.
Direct Submission
Submitted (13-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 106650)
Zhan,M. and Roe,B.A.
Direct Submission
Submitted (31-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Sep 25, 1999 this sequence version replaced gi:5919292.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below.
AC008018(b379n11) 146090 180888 (0) overlaps AC007708(b563b9) 1
34927 (71723) AC007708(b563b9) 1843 106650 (0) overlaps
AC009288(p413m7) 1 101851 (39025).
FEATURES
source
1.106650
/organism="Homo sapiens"
/db_xref="taxon:9606"

REFERENCE AUTHORS TITLE JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Homo sapiens Chromosome 22q11 BAC Clone b291 In CES-DGCR Region Unpublished
REFERENCE AUTHORS TITLE JOURNAL	2 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (07-JUL-1999) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	3 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (08-JAN-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	4 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (12-JAN-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	5 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (13-JAN-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	6 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (12-MAR-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	7 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (18-MAY-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	8 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (19-MAY-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	9 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (20-MAY-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE AUTHORS TITLE JOURNAL	10 (bases 1 to 175358) Fu,Y., Pan,H., Shaikh,T., Kurahashi,H., Emanuel,B. and Roe,B.A. Direct Submission Submitted (27-MAY-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT	On May 20, 2000 this sequence version replaced g1:7940350. Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below. AC008132(pacg9506) 4747 145670 (0) overlaps AC007981(b291) 1 14242525 (335936) AC007981(b291) 59901 175358 (0) overlaps AC007325(b48) 1 115948 (491002) AC007981(b291) 137871 175358 (0) overlaps AC008103(pacg99j1) 1 44355 (542023).
FEATURES	Location/Qualifiers 1..175358 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22q11" /clone="b291"
SOURCE	

BASE COUNT	47856 a	42018 c	40291 g	45193 t		
ORIGIN						
Query Match	68.8%; Score 344; DB 9; Length 175358;					
Best Local Similarity	84.3%; Pred. No. 2,4e-23;					
Matches 413; Conservative	0; Mismatches 70; Indels 7; Gaps 2;					
OY	18	GAGACGCTGCCCGACGGCATCGCTAATATAGAACGCCACCAAGGCATGTGCTAAATGAGGAC	77			
Db	160734	GAGGACGGCCGCCACGAGGATGCGCCAGCAGAGACGCGCCACGCGATGCGCACGAGGAGAC	160695			
OY	78	ACCACCGATGTCATCGCCACGAGGAAGCCGCCACGAGGCATCGC---CGAGAGCCCATC	134			
Db	160694	GCCGCCACGGGATCGCCACGAGGAGACGCCGCCACGAGGCATCGCCACGAGGAGCCGCC	160635			
OY	135	CAGGCGATCGCCACAGAGAGGTTGTCGAGGAGCATCGCCATTGGGGTTCGCCACAGAGGC	194			
Db	160634	CAGGCGATCGCCACAGAGAGACCGGCCACAGGAGCATCGCCACAGAGACGCCGCCAGAGGC	160575			
OY	195	ATCGCCAAATGAGAGACGCCACCCAGGGCATTCGCCAACTGGAGGCCGCTTCATCGCTCGCC	254			
Db	160574	ATCGCCAAACGAGAGACTTCGCCACAGGAATCGCCACGAGAGCGCCGCCACGCGATCGCC	160515			
OY	255	AACGGGAGCGCCGTCCTCAGCTTGCCCAACGAGGGAGCGCCGCCACGAGGCATCGCCAAACGG	314			
Db	160514	AGCCGAGAGCGCCGTCCTCAAGGGCATCGCCAAAGGAGGAGCGCCGCCACGAGGCATCGCCAAACG	160455			
OY	315	GACGCGCACCAAGGGGATGGGCAAGAGGTTCACATCCACGGGATCGCTTAACGAGAGGCC	374			
Db	160454	GACGCGCGCCCAAGGATGCGCCAGAGAGACCGCGGCCACAGGGATCGCCCAAGAGAGAGGCC	160395			
OY	375	GTCACAGGCGATCGCTAACGAGAGTGGCCGCCCAAGGGCATTCGCCAACGAGAGACGCCGCCAG	434			
Db	160394	GCCCAAGGCGATCGCCACAGAGGGCGCCGCCCAAGGCGATCGCCAAAGAGAGCGCCGCCAG	160335			
OY	435	GGAAATCGCCGAGA-----TGTCGACAAGGGGATCGCCCAACGAGAGCGCGCCACGAGGCAT	490			
Db	160334	GSCATCGCCCAAGAGAGATGCGGCCACAGGGCATCGCCCAACGAGAGCGCGCCACGAGGAT	160275			
OY	491	CGCCCAACGAG	500			
Db	160274	CGCCACGCGAG	160265			
RESULT 15						
LOCUS	AC008018	180884 bp	DNA	linear PRI 31-MAY-2000		
DEFINITION	Homo sapiens chromosome 22q11 clone b379n11, complete sequence.					
ACCESSION	AC008018					
VERSION	AC008018.20	GI:7958973				
SOURCE	HTG.					
ORGANISM	human.					
REFERENCE	Homo sapiens					
AUTHORS	Euhayvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE	1 (bases 1 to 180884)					
AUTHORS	Zhan,M. and Roe,B.A.					
REFERENCE	Homo sapiens Chromosome 22q11 BAC Clone b379n11 In BCRL2-GST Region					
AUTHORS	Unpublished					
TITLE	2 (bases 1 to 180884)					
REFERENCE	Zhan,M. and Roe,B.A.					
JOURNAL	Direct Submission					
TITLE	Submitted (10-JUL-1999) Department of Chemistry and Biochemistry,					
REFERENCE	The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,					
AUTHORS	OK 73019, USA					
JOURNAL	3 (bases 1 to 180884)					
REFERENCE	Zhan,M. and Roe,B.A.					
AUTHORS	Direct Submission					
TITLE	Submitted (01-NOV-1999) Department of Chemistry and Biochemistry,					
REFERENCE	The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,					
AUTHORS	OK 73019, USA					
JOURNAL	4 (bases 1 to 180884)					
REFERENCE	Zhan,M. and Roe,B.A.					
AUTHORS						

BASE COUNT 166 a 261 c 246 g 72 t 3 others
 ORIGIN
 Query Match 100.0%; Score 500; DB 6; Length 748;
 Best Local Similarity 100.0%; Pred. No. 1,8e-36;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCTACGCGATTGCTGAGACGCTGCCAGGCGATGCTAATGAGACGCCAGCAG 60
 |||||||
 Db 1 GAGCTACGCGATTGCTGAGACGCTGCCAGGCGATGCTAATGAGACGCCAGCAG 60
 |||||||

QY 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||
 Db 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||

QY 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||
 Db 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||

QY 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||
 Db 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||

QY 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||
 Db 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||

QY 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||
 Db 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||

QY 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||
 Db 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||

QY 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||
 Db 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||

QY 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||
 Db 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||

QY 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||
 Db 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||

QY 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||
 Db 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||

QY 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||
 Db 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||

QY 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||
 Db 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||

QY 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||
 Db 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||

QY 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||
 Db 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||

QY 481 CCCAGGCGATCGCCACAGAG 500
 |||||||
 Db 481 CCCAGGCGATCGCCACAGAG 500
 |||||||

RESULT 2
 AX201594 1201 bp DNA linear PAT 30-AUG-2001
 LOCUS
 DEFINITION Sequence 15 from Patent WO0153349.
 ACCESSION AX201594
 VERSION AX201594.1 GI:15391443
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1201)
 AUTHORS Stockert, E., Scanlan, M.J., Jager, D., Old, L.J., Gure, A.O. and
 Chen, Y.T.
 TITLE Small cell lung cancer associated antigens and uses therefor
 PATENT: WO 0153349-A 15 26-JUL-2001.
 JOURNAL LUDWIG INSTITUTE FOR CANCER RESEARCH (US); MEMORIAL
 SLOAN-KETERING CANCER CENTER (US); CORNELL RESEARCH FOUNDATION,
 INC. (US)
 FEATURES
 Location/Qualifiers
 source 1..1201
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 ORIGIN 255 a 377 c 349 g 178 t 42 others

QY 1 GAGCTACGCGATTGCTGAGACGCTGCCAGGCGATGCTAATGAGACGCCAGCAG 60
 |||||||
 Db 1 GAGCTACGCGATTGCTGAGACGCTGCCAGGCGATGCTAATGAGACGCCAGCAG 60
 |||||||

QY 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||
 Db 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||

QY 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||
 Db 61 GCATCGCTAATGAGACACCCAGTGCGATCGCCACAGAGAACGCCAGGCGATCG 120
 |||||||

QY 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||
 Db 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||

QY 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||
 Db 121 CCGAGAGCGCATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 180
 |||||||

QY 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||
 Db 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||

QY 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||
 Db 181 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 240
 |||||||

QY 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||
 Db 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||

QY 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||
 Db 241 TCAGCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 300
 |||||||

QY 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||
 Db 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||

QY 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||
 Db 301 GCATCGCGATTCAGGCGATCGCCACAGAGAGTTGCCAGGCGATCGCCAGTGGG 360
 |||||||

QY 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||
 Db 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||

QY 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||
 Db 361 CTACAGAGAGAGCGCGCTCCAGAGGATGCTTAACAGAGTGGCGCCAGGCGATCGCCAGC 420
 |||||||

QY 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||
 Db 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||

QY 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||
 Db 421 AGAGCGCGCGCCAGGAGATCGCCAGAGATGTCGACAGAGGCGATCGCCAGAGAGCGCG 480
 |||||||

QY 481 CCCAGGCGATCGCCACAGAG 500
 |||||||
 Db 481 CCCAGGCGATCGCCACAGAG 500
 |||||||

RESULT 3
 HSM801408 2448 bp mRNA linear PRI 18-FEB-2000
 LOCUS
 DEFINITION Homo sapiens mRNA; cDNA DKF2p434C196 (from clone DKF2p434C196);
 partial cds.
 ACCESSION AL133561
 VERSION AL133561.1 GI:6599133
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2448)
 AUTHORS Koehn, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 DIRECT SUBMISSION
 Submitted (15-DEC-1999) MIPS, Am Klopfersplitz 18a, D-82152
 Martinsried, GERMANY
 CLONE FROM S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by BMFZ (Biomedical Research Center at the Charité,
 Berlin/Germany) within the cDNA sequencing consortium of the German
 Genome Project.
 This clone (DKFZ434C196) is available at the RZPD in Berlin.
 Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cdnh/>.
 FEATURES
 Location/Qualifiers
 source 1..2448
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKF2p434C196"
 /rname="DKF2p434C196"
 /rsize="434"
 /rtype="mRNA"
 /rclone="434 (synonym: htes3). Vector pSport1; host
 DH10B; sites NotI + SalI"


```

repeat_region /evidence-not_experimental
11054. .11343 /rpt_family="AluSx"
/evidence-not_experimental
repeat_region 11344. .11367 /rpt_family="(TTAA)n"
/evidence-not_experimental
repeat_region complement(12418. .12736)
/rpt_family="AluSg"
/evidence-not_experimental
repeat_region complement(13993. .14328)
/rpt_family="AluJo"
/evidence-not_experimental
repeat_region complement(14759. .14875)
/rpt_family="FLAM_C"
/evidence-not_experimental
repeat_region 14876. .14886 /rpt_family="Charlie4"
/evidence-not_experimental
repeat_region complement(14687. .15155)
/rpt_family="AluSx"
/evidence-not_experimental
repeat_region 15156. .15239 /rpt_family="Charlie4"
/evidence-not_experimental
repeat_region 15618. .15788 /rpt_family="L2"
/evidence-not_experimental
repeat_region 15911. .15924 /rpt_family="L1MB7"
/evidence-not_experimental
repeat_region complement(15925. .16230)
/rpt_family="AluJo"
/evidence-not_experimental
repeat_region 16231. .16940 /rpt_family="L1MB7"
/evidence-not_experimental
repeat_region complement(17656. .17816)
/rpt_family="FLAM_A"
/evidence-not_experimental
repeat_region complement(17861. .18018)
/rpt_family="AluJo"
/evidence-not_experimental
repeat_region 18019. .18049 /rpt_family="(TTTA)n"
/evidence-not_experimental
repeat_region complement(18050. .18176)
/rpt_family="AluJo"
/evidence-not_experimental
repeat_region 18300. .18602 /rpt_family="AluJo"
/evidence-not_experimental
repeat_region complement(19466. .19612)
/rpt_family="L2"
/evidence-not_experimental
repeat_region 19721. .20022 /rpt_family="AluY"
/evidence-not_experimental
repeat_region complement(20025. .20302)
/rpt_family="AluSg"
/evidence-not_experimental
repeat_region complement(20308. .20325)
/rpt_family="AluSp/q"
/evidence-not_experimental
repeat_region 20326. .20374 /rpt_family="(TG)n"
/evidence-not_experimental
repeat_region complement(20375. .20445)
/rpt_family="AluSp/q"
/evidence-not_experimental
repeat_region 20593. .20669 /rpt_family="AluSp/q"
/evidence-not_experimental

```

```

repeat_region complement(20885. .21149)
/rpt_family="AluY"
/evidence-not_experimental
repeat_region 21186. .21481 /rpt_family="AluY"
/evidence-not_experimental
repeat_region complement(21482. .21758)
/rpt_family="AluSg"
/evidence-not_experimental
repeat_region complement(21759. .21862)
/rpt_family="AluSp/q"
/evidence-not_experimental
repeat_region 24439. .24588 /rpt_family="AluY"
/evidence-not_experimental
repeat_region complement(24678. .24968)
/rpt_family="AluSx"
/evidence-not_experimental
repeat_region complement(25016. .25306)
/rpt_family="AluYb8"
/evidence-not_experimental
repeat_region complement(25310. .25609)
/rpt_family="AluYb8"
/evidence-not_experimental
repeat_region 25627. .25853 /rpt_family="AluSx"
/evidence-not_experimental
repeat_region 25859. .25885 /rpt_family="(A)n"
/evidence-not_experimental
repeat_region 26308. .26588 /rpt_family="AluJo"
/evidence-not_experimental
repeat_region complement(30748. .31041)
/rpt_family="AluJo"
/evidence-not_experimental
repeat_region 35559. .35661 /rpt_family="AluSg/x"
/evidence-not_experimental
repeat_region complement(35853. .35996)
/rpt_family="MER5B"
/evidence-not_experimental
repeat_region 36421. .36721 /rpt_family="AluSg"
/evidence-not_experimental
repeat_region 37016. .37083 /rpt_family="MUT1E"
/evidence-not_experimental
repeat_region 37175. .37372 /rpt_family="MUT1E2"
/evidence-not_experimental
repeat_region 37480. .37793 /rpt_family="AluY"
/evidence-not_experimental
repeat_region 38069. .38351 /rpt_family="AluSg"
/evidence-not_experimental
repeat_region 38506. .38612 /rpt_family="MIR"
/evidence-not_experimental
repeat_region 38841. .38985 /rpt_family="AluSg/x"
/evidence-not_experimental
repeat_region 38987. .39118 /rpt_family="L1PA16"
/evidence-not_experimental
repeat_region 39119. .39173 /rpt_family="L1PA16"

```

Query Match 71.9%; Score 359.4; DB 9; Length 157086;
 Best Local Similarity 86.3%; Pred. No. 1.1e-24;
 Matches 422; Conservative 0; Mismatches 61; Indels 6; Gaps 2;

Oy	18	GAGGAGCCTGCTCCCAAGGGCATTCGCTATATGAGAGCGCCGACACAGGGCATTCGCTAAATGAGAC	77
Db	48981	GAGGAGCCTCCCAAGGGCATTCGCTAAATGAGAGCGCCGCTCCACAGGGCATTCGCTAAAGAGAC	48922
Oy	78	ACCCACCCATGTCATTCGCCCAAGAGAGAGACGCCGCCACAGGGCATTCGCCGCGGAGAGGCCATC	134
Db	48921	GCCCCCAAGGGCATTCGCCCAAGAGAGAGCGCCGCCGCCAGGGCATTCGCCCAAGAGAGAGACGCCGCC	48862
Oy	135	CAGGGCATTCGCCCAAGAGAGAGGTGGCCCGAGGGCATTCGCCCAATGGAGTGGGTCGCCGCCACAGGG	194
Db	48861	CAGGGCATTCGCCCAAGAGAGAGCGCCGCCGCCAGGGCATTCGCCCAAGAGAGAGACGCCGCCACAGGG	48802
Oy	195	ATTCGCATATGAGAGAGCGCACCCAGGGCATTCGCCCAATCGTGGAGAGCGCTCCACAGGCTTCGCC	254
Db	48801	ATTCGCATAGAGAGAGCGCGCCGCCCGAGGGCATTCGCCCAAGAGAGAGCGCGCCGCCCGAGGGCATTCGCC	48742
Oy	255	AACGGGAGCGCGCTTCCTCAGCTTCGCCCAAGGGAGAGCGCGGCCCGAGGGCATTCGCCCAAGGG	314
Db	48741	AAAGAGAGCGCGCCCATGTGTCATTCGCCCAAGAGAGAGCGCGGCCCGAGGGCATTCGCCCAAGAG	48682
Oy	315	GAGCCACACAAAGGGCATATGGGACAAAGAGTCATCATTCACAGGCATTCGCTAAACAGAGAGCGC	374
Db	48681	GAGCCCGCCCAAGGATTCGCCACAGAGAGAGCGCGGCCCGAGGGCATTCGCCCAAGAGAGAGCGCC	48622
Oy	375	GTCACAGGGCATTCGCTTAACAGGTGGCGCCGCCAGGGCATTCGCCCAAGAGAGAGCGCGCCCGAG	434
Db	48621	GTCACAGGGCATTCGCCCAAGAGAGAGTGGCGCGGCCCGAGGGCATTCGCCCAAGAGAGAGCGCGCCCGAG	48562
Oy	435	GGAATGCGCGCGAGATGTCGACACAGGGCATTCGCCCAAGAGAGAGCGCGCCCGAGGGCATTC	491
Db	48561	GGAATGTCGCACACAGGACGCGCGGCCCGAGGGCATTCGCCCAAGAGAGAGCGCGCCCGAGGGCATTC	48502
Oy	492	GCCACACAG	500
Db	48501	GCCACACAG	48493

RESULT	b
LOCUS	AC023491/c
DEFINITION	Homo sapiens chromosome 22q11 clone unknown, complete sequence.
VERSION	AC023491
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 150754)
JOURNAL	Hu,P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
REFERENCE	Humo sapiens Chromosome 22q11 BAC Clone 659m11 In BCRU2-GGT Region
AUTHORS	Unpublished
TITLE	2 (bases 1 to 150754)
JOURNAL	Hu,P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
REFERENCE	Direct Submission
AUTHORS	Submitted (14-FEB-2000) Department Of Chemistry And Biochemistry,
TITLE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
JOURNAL	OK 73019, USA
REFERENCE	3 (bases 1 to 150754)
AUTHORS	Hu,P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Department Of Chemistry And Biochemistry,
REFERENCE	OK 73019, USA
AUTHORS	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
TITLE	OK 73019, USA
JOURNAL	4 (bases 1 to 150754)
REFERENCE	Hu,P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
AUTHORS	Direct Submission
TITLE	Submitted (01-JUL-2000) Department Of Chemistry And Biochemistry,
JOURNAL	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
REFERENCE	OK 73019, USA
AUTHORS	5 (bases 1 to 150754)
TITLE	Hu,P., Wu,H., Yang,L., Morrow,B.E. and Roe,B.A.
JOURNAL	Direct Submission

JOURNAL	REFERENCE AUTHORS TITLE	JOURNAL	COMMENT FEATURES	source
Submitted (20-JUL-2000) Department of Chemistry, And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	6 (bases 1 to 150754) Hu, P., Wu, H., Yang, L., Morrow, B.E. and Roe, B.A. Direct Submission	Submitted (21-JUL-2000) Department of Chemistry, And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	On Jul 21, 2000 this sequence version replaced gi:9295769. location/Qualifiers 1..150754 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="2q31.1" /clone="unknown" /clone_lib="unknown"	38247 a 34915 c 37024 g 40568 t

	Query Match	71.6%;	Score 357.8;	DB 9;	Length 150754;
	Best Local Similarity	86.1%;	Pred. No. 1.5e-24;		
	Matches 421;	Conservative	0;	Mismatches 62;	Indels 6;
				Gaps	2;
QY 18	GAGGAGCCTCCCGAGGCGATCGCTATGTAGGAGCGCCGACACAGGGCGATCGCTAATGAGAGAC	77			
Db 111233	GAGGAGCGCCGCCGAGGGCATTCGCGAAGCAGAGACGCCGCCCGCAGGGCGATCGCGAAGAGAGAC	111174			
QY 78	ACACCCCGATGCTATCGCCCAAGCAGAGAGAGCGCCGCCAGGGCGATCGCC---GAGAGCGCATC	134			
Db 111173	GCGGCCGAGGGCATTCGCCCAAGAGAGAGCGCCGCCCGCAGGGCGATTCGCCCAAGAGAGAGCGCCGC	111114			
QY 135	CAGGGCATTCGCCCAAGCAGAGAGATTGCCCGCAGGGCGATTCGCCCAATGGGATCGCCCGACAGAGGC	194			
Db 111113	CAGGGCATTCGCCCAAGCAGAGAGAGCGCGCCCGCAGGAAATGCGCAACAGAGAGAGCGCGCCCGCAGGGC	111054			
QY 195	ATCGCCATGAGGAGCGCCACCCAGGAGCGCATTCGCCCAACTGGGAGCGCGCTTCACAGCGCTTCGCC	254			
Db 111053	ATCGCCCAACGAGAGAGCGCCCGCCCGAGGGCATTCGCCACGAGAGAGCGCGCCCGCAGGGCGATTCGAC	110994			
QY 255	AACGGGAGCGCCGCTCTTCAGCTTCGCCCAACGGGGAGCGCGCGCCCGCAGGGCGATTCGCCCAACGGG	314			
Db 110993	AAGGAGAGCGCCCGCCCGCAGGGCATTCGCCCAACGAGAGCGCGCCCGCAGGGCGATTCGCCCAACGAG	110934			
QY 315	GAGCGCACCAAGGGGCTATGGGGCAAGCAGGTACATCCACGGCGATTCGGTTAACGAGAGAGAGCGCC	374			
Db 110933	GAGCGCGCCCGCAGGGCATTCGCCCAACGAGAGCGCGCATTCGCCCAAGGGCGATTCGCCCAAGAGAGTACG	110874			
QY 375	GTCGAGGGCATTCGCTTAACGAGGTGGCGCGCCCGCAGGGCGATTCGCCCAACGAGAGAGAGCGCGCCAG	434			
Db 110873	GTCGAGGGCATTCGCCCAACGAGAGAGCGCGCCCGCAGGGCGATTCGCCCAAGAGAGAGAGCGCGCCAG	110814			
QY 435	GGATATGCC---GAGATGTTCGCGACAGGGCGATTCGCCCAACGAGAGAGCGCGCCCGCAGGGCGATC	491			
Db 110813	GGATATGCCCAAGGAGAGAGCGCGCCCGCAGGGCGATTCGCCCAAGAGAGAGAGCGCGCCCGCAGGGCGATC	110754			
QY 492	GCCACACGAG 500				
Db 110753	GCCACACGAG 110745				
RESULT 7	AP000550	150724 bp	DNA	linear	PRI 01-OCT-1999
LOCUS	AP000550	150724 bp	DNA	linear	PRI 01-OCT-1999
DEFINITION	AP000550	150724 bp	DNA	linear	PRI 01-OCT-1999
ACCESSION	AP000550	150724 bp	DNA	linear	PRI 01-OCT-1999
VERSION	AP000550	150724 bp	DNA	linear	PRI 01-OCT-1999
KEYWORDS	AP000550.1	GT:5931536			
SOURCE	Homo sapiens pre-pro-B cell	clone:Kb1592A4.			
ORGANISM	Homo sapiens	BAC library clone:Kb1592A4.			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

```

REFERENCE 1 (bases 1 to 150724)
AUTHORS Shimizu,N.
TITLE Human DNA sequence from clone KB1592A4 on chromosome 22q11.2
JOURNAL Published Only in Database (1999) In press
REFERENCE 2 (bases 1 to 150724)
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1999) to the DDBJ/EMBL/GenBank databases.
Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan
Fax:81-3-3351-2370
(E-mail:nshimizu@med.keio.ac.jp, Tel:81-3-3351-2370,
This is a complete sequence of the insert of KB1592A4 clone. The
distal adjacent clone is KB876E2 (Acc.#AF005511) with 1196-bp
overlapping.

COMMENT
FEATURES
    source
        1..150724
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="FEB14-14"
            /cell_type="pre-pro-B cell"
            /chromosome="22"
            /clone="KB1592A4"
            /map="22q11.2"
            /clone_id="Keio BAC library"
            /complement(1..90)
            /rpt_family="Alus"
            /evidence-not_experimental
            /complement(99..399)
            /rpt_family="Alusg"
            /evidence-not_experimental
            /complement(401..570)
            /rpt_family="L1R33"
            /evidence-not_experimental
            /complement(572..640)
            /rpt_family="MIR"
            /evidence-not_experimental
            /complement(649..776)
            /rpt_family="TCCA)n"
            /evidence-not_experimental
            /complement(717..752)
            /rpt_family="L2"
            /evidence-not_experimental
            /complement(757..862)
            /rpt_family="MIR"
            /evidence-not_experimental
            /complement(933..1059)
            /rpt_family="L2"
            /evidence-not_experimental
            /complement(1955..2151)
            /rpt_family="Aluub"
            /evidence-not_experimental
            /complement(2578..2875)
            /rpt_family="Alusx"
            /evidence-not_experimental
            /complement(3310..3408)
            /rpt_family="L1R16A"
            /evidence-not_experimental
            /complement(4424..4591)
            /rpt_family="L1PA4"
            /evidence-not_experimental
            /complement(4617..4637)
            /rpt_family="(TAAA)n"
            /evidence-not_experimental
            /complement(4963..5286)
            /rpt_family="L1"
            /evidence-not_experimental
            /complement(5353..5554)
            /rpt_family="L1"
            /evidence-not_experimental
            /complement(5542..5625)
            /rpt_family="L1M1"

repeat_region /evidence-not_experimental
complement(5626..5882)
rpt_family="Aluub"
/evidence-not_experimental
complement(5883..6008)
rpt_family="L1M1"
/evidence-not_experimental
complement(6025..6235)
rpt_family="Aluub"
/evidence-not_experimental
complement(6236..6447)
rpt_family="L1M2"
/evidence-not_experimental
complement(6486..6554)
rpt_family="Alu/FRAM"
/evidence-not_experimental
complement(6558..6709)
rpt_family="L1"
/evidence-not_experimental
complement(6778..7721)
rpt_family="L1M6"
/evidence-not_experimental
complement(8075..8296)
rpt_family="L1PA8"
/evidence-not_experimental
complement(8679..8980)
rpt_family="Alusx"
/evidence-not_experimental
complement(9224..9298)
rpt_family="L1M2"
/evidence-not_experimental
complement(9324..9932)
rpt_family="L1M2"
/evidence-not_experimental
complement(10013..10508)
rpt_family="L1M2"
/evidence-not_experimental
complement(10548..10734)
rpt_family="MIR"
/evidence-not_experimental
complement(10867..11330)
rpt_family="L1MA7"
/evidence-not_experimental
complement(11332..11537)
rpt_family="L1M2"
/evidence-not_experimental
complement(11540..11728)
rpt_family="Alusg/x"
/evidence-not_experimental
complement(11730..11910)
rpt_family="L1M2"
/evidence-not_experimental
complement(11911..12230)
rpt_family="L1PA6"
/evidence-not_experimental
complement(12231..12416)
rpt_family="L1P"
/evidence-not_experimental
complement(12640..12746)
rpt_family="MIR"
/evidence-not_experimental
complement(12901..13183)
rpt_family="Alusg"
/evidence-not_experimental
complement(13459..13774)
rpt_family="Aluy"
/evidence-not_experimental
complement(13882..14031)
rpt_family="MTR1E2"
/evidence-not_experimental
complement(14171..14238)
rpt_family="MTR1E"
/evidence-not_experimental

```

[illegible]

OY	375	GTAAGGGCATGCTTACGAGGTGGCGGCCGCACGAGGATTCGCAACAAGAGAAGCCGCCCGCAG	434
Db	84991	CCCCAGGGCATTCGCCAACAGAGACC GCCGCCGACGAGGCATCGCCAAACGAAGACGCCGCCAG	85050
OY	435	GGATTGCGC---GAGATGTGCGACANGGCATTCGCCAACAAGAGAAGCCGCCGCAGGCATTC	491
Db	85051	GGCATTCGCCAACAAGAGACCCGCCGAGGCATTCGCCAAGAAGAGACGCCGCCGAGGCATTC	85110
OY	492	CCCAAACAAG 500 	
Db	85111	CCCAAACGAG 85119	
RESULT	8		
LOCUS	AC007324	122364 bp	DNA linear PRI 18-MAY-2000
DEFINITION	Homo sapiens chromosome 22q11 clone b293,	complete sequence.	
ACCESSION	AC007324		
VERSION	AC007324.55	GI:7923342	
KEYWORDS	HYG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 122364)		
JOURNAL	Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A. Homo sapiens Chromosome 22q11 BAC Clone b293 In CES Region Unpublished		
REFERENCE	2 (bases 1 to 122364) Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A. Direct Submission Submitted (17-Apr-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
AUTHORS	3 (bases 1 to 122364) Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A. Direct Submission Submitted (19-Dec-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman, OK 73019, USA		
REFERENCE	4 (bases 1 to 122364) Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A. Direct Submission Submitted (06-May-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
AUTHORS	5 (bases 1 to 122364) Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A. Direct Submission Submitted (18-May-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
COMMENT	On May 18, 2000 this sequence version replaced gi:7712128.		
FEATURES	Location/Qualifiers		
SOURCE	1..122364 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22q11" /clone="b293"		
BASE COUNT	31795 a 28253 c 28482 g 33834 t		
ORIGIN			
Query Match	70.9%:	Score 354.6:	DB_9: Length 122364:
Best Local Similarity	85.7%:	Pred. No. 3,1e-24:	
Matches 419;	Conservative 0;	Mismatches 64;	Indels 6;
		Gaps 2;	
OY	18	GAGAGACCTGCCACGAGGCATCTCTAATAGNAGCCGCCACAGAGGCATCGTTATGAGAGC	77
Db	82628	GAGAGACCTGCCACGAGGCATCTCTAATAGNAGCCGCCACAGAGGCATCGTTATGAGAGC	82687
OY	78	ACCAACCCATGTGATCGCCAACAAGAGAGCGCCGACGAGGCATCGC--CGAGAGCCGATTC	134
Db	82688	GCCTCCCACAGGGCATTCGCCAACAAGAGAGCGCCGACGAGGCATCGCCAACAAGAGCGCGCGCC	82747

```

QY 135 CAGGCGATCGCCACGAGAGGTTGCCAGGCGCATCGCCATGGGTCGCCGACGAGGC 194
DB 82748 CAGGCGATCGCCACGAGAGGAGCGCCGCCAGGCGATCGCCACGAGAGGCGCCGACGAGGC 82807
QY 195 ATTCGCAATGAGGAGAGCGCCAGGCGCATCGCCACACTGGGAGCGCGTCACGCGCTTCGCC 254
DB 82808 ATTCGCAATGAGGAGAGCGCGCCAGGCGCATCGCCACACTGGGAGCGCGTCACGCGCTTCGCC 82867
QY 255 AACGGGAGCGCGCTTCACGCTTCGCAAGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
DB 82868 AACGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82927
QY 315 GACGCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 374
DB 82928 GACGCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82987
QY 375 GTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
DB 82988 GTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 83047
QY 435 GGAATCGCC--GAGGATGTCGCGCAGAGGCGCATCGCCACGAGAGCGCGCGCGCGCGCGCGCG 491
DB 83048 GGAATCGCC--GAGGATGTCGCGCAGAGGCGCATCGCCACGAGAGCGCGCGCGCGCGCGCGCG 83107
QY 492 GCCACACAG 500
DB 83108 GCCACACAG 83116

RESULT 9
AC013360 159550 bp DNA linear HTG 04-MAY-2001
LOCUS Homo sapiens chromosome 11, clone RP11-278E23 map 11, WORKING DRAFT
DEFINITION AC013360.13
ACCESSION AC013360.4 GI:7229982
VERSION AC013360.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 159550)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Homo sapiens chromosome 11, clone RP11-278E23
REFERENCE 2 (bases 1 to 159550)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguski, L., Bouhassira, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, S., Collins, S.,
Cooke, P., Deari, K., Desai, K., Dominko, T., Donnelly, L., Doyle, M.,
Farrington, P., Fitzhugh, W., Forrest, C., Funk, R., Gage, D., Gage, D.,
Gallagher, J., Gargano, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
Howard, J., Johnson, R., Jones, C., Kahn, L., Karas, A., Klein, D.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McGowan, P., McGurk, A., McKernan, K., McDonald, J., Meldrum, J.,
Morrow, J., Nayler, J., Norman, C., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santol, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testa, S., Tjelle, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 330 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced by: 6479088.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```

----- Project Information
Center project name: L2604
Center clone name: 278_E_23
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960721
Consensus quality: 144663 bases at least Q40
Consensus quality: 153318 bases at least Q30
Insert size: 173000; agarose-fp
Insert size: 158350; sum-of-ctrls
Quality coverage: 5.4 in Q20 bases; agarose-fp
Quality coverage: 5.9 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1729: contig of 1729 bp in length
1730 1829: gap of 100 bp
1830 3716: contig of 1887 bp in length
3717 3816: gap of 100 bp
3817 6950: contig of 3134 bp in length
6951 7050: gap of 100 bp
7051 9136: contig of 2086 bp in length
9137 9236: gap of 100 bp
9237 17859: contig of 8623 bp in length
17860 17959: gap of 100 bp
17960 26939: contig of 8980 bp in length
26940 27039: gap of 100 bp
27040 36269: contig of 9230 bp in length
36270 36369: gap of 100 bp
36370 48574: contig of 12205 bp in length
48575 48674: gap of 100 bp
48675 62315: contig of 13641 bp in length
62316 62415: gap of 100 bp
62416 77457: contig of 15042 bp in length
77458 77557: gap of 100 bp
77558 96884: contig of 19327 bp in length
96885 96984: gap of 100 bp
96985 124230: contig of 27246 bp in length
124231 124330: gap of 100 bp
124331 159550: contig of 35220 bp in length.
-----
FEATURES
source
1. 159550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-278E23"
/clone_id="RP11-11 Human Male BAC"
1. 1729
/note="assembly-fragment"
1830. 3716
/note="assembly-fragment"
3817. 6950
/note="assembly-fragment"
7051. 9136
/note="assembly-fragment"
9237. 17859
/note="assembly-fragment"
17960. 26939
/note="assembly-fragment"
27040. 36269
/note="assembly-fragment"
36370. 48574
/note="assembly-fragment"
48675. 62315

```



```

BASE COUNT      30841 a 24372 c 23166 g 28271 t
ORIGIN
Query Match      69.3%; Score 346.6; DB 9; Length 106550;
Best Local Similarity 84.7%; Pred. No. 1.7e-23;
Matches 414; Conservative 0; Mismatches 69; Indels 6; Gaps 2;

18 GAGAGCGTCCCGCAGGCGATCGCTAATGAGACGCCAGCGGCGATCGCTAATGAGAC 77
DB 80846 GAGAGCGTCCCGCAGGCGATCGCTAATGAGACGCCAGCGGCGATCGCTAATGAGAC 80787
QY 78 ACACCGCATGTCATCGCCAGGAGAGAGCGCGCCAGGCGATCGCTAATGAGAC 134
DB 80786 GCGCCCGCAGGCGATCGCCAGGAGAGAGCGCGCCAGGCGATCGCTAATGAGAC 80727
QY 135 CAGGCGATCGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
DB 80726 CAGGCGATCGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80667
QY 195 ATCCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
DB 80666 ATCCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80607
QY 255 AACGGGGAGCGCGCTTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
DB 80606 AACGGGGAGCGCGCTTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80547
QY 315 GAGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 374
DB 80546 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80487
QY 375 GTCCAGGCGATCGCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 434
DB 80486 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80427
QY 435 GGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 491
DB 80426 GCGATCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80367
QY 492 GCCAACAAG 500
DB 80366 GCCAACAAG 80358

RESULT 13
AC009288/c 140876 bp DNA linear PRI 31-MAY-2000
LOCUS Homo sapiens chromosome 22q11 clone p413m7, complete sequence.
DEFINITION AC009288
ACCESSION AC009288.13 GI:7798764
VERSION HTG.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hua, A., Emanuel, B. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 140876)
Hua, A., Emanuel, B. and Roe, B. A.
REFERENCE
AUTHORS Hua, A., Emanuel, B. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 140876)
Hua, A., Emanuel, B. and Roe, B. A.
REFERENCE
AUTHORS Hua, A., Emanuel, B. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 140876)
Hua, A., Emanuel, B. and Roe, B. A.

```

```

REFERENCE
AUTHORS Hua, A., Emanuel, B. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 140876)
Hua, A., Emanuel, B. and Roe, B. A.
REFERENCE
AUTHORS Hua, A., Emanuel, B. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On May 13, 2000 this sequence version replaced g1:6056212.
FEATURES
Location/Qualifiers
1..140876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q11"
/clone="p413m7"
BASE COUNT      38395 a 33854 c 32663 g 35964 t
ORIGIN
Query Match      69.0%; Score 345; DB 9; Length 140876;
Best Local Similarity 84.5%; Pred. No. 2.1e-23;
Matches 413; Conservative 0; Mismatches 70; Indels 6; Gaps 2;

18 GAGAGCGTCCCGCAGGCGATCGCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77
DB 78832 GAGAGTGCCTCCCGCAGGCGATCGCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78773
QY 78 ACCAGCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 134
DB 78772 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78713
QY 135 CAGGCGATCGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
DB 78712 AAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78653
QY 195 ATCCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
DB 78652 ATCCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78593
QY 255 AACGGGGAGCGCGCTTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
DB 78592 AACGGGGAGCGCGCTTCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78533
QY 315 GAGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 374
DB 78532 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78473
QY 375 GTCCAGGCGATCGCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 434
DB 78472 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78413
QY 435 GGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 491
DB 78412 GCGATCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78353
QY 492 GCCAACAAG 500
DB 78352 GCCAACAAG 78344

RESULT 14
AC007981/c 175358 bp DNA linear PRI 27-MAY-2000
LOCUS Homo sapiens chromosome 22q11 clone D291, complete sequence.
DEFINITION AC007981
ACCESSION AC007981.46 GI:7958972
VERSION HTG.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


C	1	103.2	21.0	543	6	5273901-6	Patent No. 5273901
C	2	104.2	20.8	1656	3	US-08-655-188-17	Sequence 17, Appl
C	3	104.2	20.8	1656	3	US-08-655-227-17	Sequence 17, Appl
C	4	104.2	20.8	1656	3	US-08-655-241-17	Sequence 17, Appl
C	5	104.2	20.8	1656	4	US-09-398-326-17	Sequence 17, Appl
C	6	104.2	20.8	2241	4	US-09-144-759-17	Sequence 17, Appl
C	7	104.2	20.8	2295	3	US-09-144-759-19	Sequence 17, Appl
C	8	104.2	20.8	2301	3	US-09-144-759-21	Sequence 21, Appl
C	9	102.6	20.5	2793	1	US-08-209-747-1	Sequence 21, Appl
C	10	102.6	20.5	2793	1	US-08-458-298-1	Sequence 1, Appl
C	11	102.6	20.5	2970	5	PCT-US92-06391-1	Sequence 1, Appl
C	12	101.8	20.4	688	4	US-08-998-416-95	Sequence 1, Appl
C	13	100.4	20.1	320	4	US-09-165-264-7	Sequence 15, Appl
C	14	100	20.0	319	4	US-09-165-264-8	Sequence 7, Appl
C	15	99.4	19.9	320	4	US-09-165-264-11	Sequence 8, Appl
C	16	99.4	19.9	320	4	US-09-165-264-13	Sequence 11, Appl
C	17	99.8	19.8	318	4	US-09-165-264-12	Sequence 13, Appl
C	18	99	19.8	2790	2	US-08-176-661-1	Sequence 12, Appl
C	19	98.4	19.7	320	4	US-09-165-264-14	Sequence 1, Appl
C	20	98.4	19.7	320	4	US-09-253-691-3	Sequence 14, Appl
C	21	98.4	19.7	3489	2	US-08-728-323A-1	Sequence 3, Appl
C	22	98.4	19.7	32207	2	US-08-770-379-20	Sequence 1, Appl
C	23	98.4	19.7	32207	4	US-08-757-669A-20	Sequence 20, Appl
C	24	98.4	19.7	32207	4	US-09-230-371A-20	Sequence 20, Appl
C	25	97.4	19.5	477	4	US-09-135-994-1	Sequence 1, Appl
C	26	92.6	18.5	533	6	5488709-5	Sequence 1, Appl
C	27	91.2	18.2	234	1	US-08-469-802B-3	Patent No. 5487705

C 28	59.2	18.2	234	2	US-08-267-803B-3	Sequence 3, Appl1
C 29	50.8	18.2	2165	2	US-08-069-721-2	Sequence 2, Appl1
C 30	90.8	18.2	2165	4	US-09-189-344-2	Sequence 2, Appl1
C 31	90.8	18.2	6530	2	US-08-146-930-1	Sequence 1, Appl1
C 32	90.8	18.2	6530	3	US-08-458-240-1	Sequence 1, Appl1
C 33	90.8	18.2	6530	1	PCT-US93-03993-1	Sequence 1, Appl1
C 34	90.2	18.0	9531	1	US-08-036-200-93	Sequence 93, Appl
C 35	90.2	18.0	9551	2	US-08-000-644-93	Sequence 93, Appl
C 36	85.4	17.1	2190	4	US-09-025-168-19	Sequence 19, Appl
C 37	84.8	17.0	6192	2	US-08-479-537A-1	Sequence 1, Appl1
C 38	84.8	17.0	6192	4	US-09-083-116-1	Sequence 1, Appl1
C 39	84.8	17.0	6192	4	US-09-134-916A-1	Sequence 1, Appl1
C 40	84.8	17.0	6449	4	US-08-479-537A-4	Sequence 4, Appl1
C 41	84.8	17.0	6449	4	US-09-083-116-4	Sequence 4, Appl1
C 42	84.8	17.0	6449	4	US-09-134-916A-4	Sequence 4, Appl1
C 43	84.4	16.9	18596	4	US-09-118-448-11	Sequence 11, Appl
C 44	84.4	16.9	15231	3	US-09-128-155-16	Sequence 16, Appl
C 45	84.2	16.9	15202	3	US-08-922-633-21	Sequence 21, Appl

ALIGNMENTS

```

RESULT 1
5273901-6/C
Patent No. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUSTUSIEN, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
SPOROZOITE 21.5 KB ANTIGEN, AC-6B
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/581,693
FILING DATE: 12-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1988
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO:6:
LENGTH:543
5273901-6

```

21.0%; Score 105.2; DB 6; Length 543;

```

Best Local Similarity  30.36;  Prev. NO.  4.1e-14;
Matches  188;  Conservative  0;  Mismatches  146;  Indels  0;  Gaps  0;

```

[illegible]

Fri Oct 11 09:30:52 2002

us-09-489-101a-17_copy_1_500.rni

Page 2

```

RESULT 2
US-08-659-188-17/c
; Sequence 17, Application us/08659188
; Patent No. 6002069
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Seed Plants Exhibiting Inducible Early
; TITLE OF INVENTION: Reproductive Development and Methods of Making Same
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,188
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 1946
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1651
; NAME/KEY: misc_feature
; LOCATION: 1..1656
; OTHER INFORMATION: /note= "domain = ecdysone receptor
; OTHER INFORMATION: ligand binding domain."
US-08-659-188-17

Query Match          20.8%; Score 104.2; DB 3; Length 1656;
Best Local Similarity 67.8%; Pred. No. 8.2e-14;
Matches 145; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 10 GGGCTGGGGCTGGGGCTGGAGCTGAGCTGAGCTGAGGCTGGGGCTGGGGCTGG 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1341 GGGCACGGAGCGAGACGAGAGAGAGCTGTGGCTGTGGTGAATCTGTGGCTG 1282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 GGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1281 CGTCTGAAGCTGTGTGTGAGCTGGGGCTGGAGCTTGACCTTGACCTG 1222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 AGGCTGGGGCTGGAGCTGAGGCTGGGGCTGGAGCTGAGGCTGGGGCTGG 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1221 AGGTTGAGCTGGCGCTGTGTCTGTGGGGAATCTTCTGGGTCAAGGAGG 1162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 GGGCTGGGACTGAGGCTGGGGCTTANGGCTGGGGC 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1161 GGGCTGGGCTGAGGCTGAGGCTGATGCTGGGCC 1128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
US-08-655-227-17/c
?
```

```

; Sequence 17, Application us/08655227
; Patent No. 6025483
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; TITLE OF INVENTION: Maize and Cauliflower APEATA1 Gene
; TITLE OF INVENTION: Products and Nucleic Acid Molecules Encoding Same
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,227
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UD 2143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1651
; NAME/KEY: misc_feature
; LOCATION: 1..1656
; OTHER INFORMATION: /note= "domain = ecdysone receptor
; OTHER INFORMATION: ligand binding domain."
US-08-655-227-17

Query Match          20.8%; Score 104.2; DB 3; Length 1656;
Best Local Similarity 67.8%; Pred. No. 8.2e-14;
Matches 145; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 10 GGGCTGGGGCTGGGGCTGGAGCTGAGCTGAGCTGAGGCTGGGGCTGGGGCTGG 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1341 GGGCACGGAGCGAGACGAGAGAGAGCTGTGGCTGTGGTGAATCTGTGGCTG 1282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 GGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1281 CGTCTGAAGCTGTGTGTGAGCTGGGGCTGGAGCTTGACCTTGACCTG 1222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 AGGCTGGGGCTGGAGCTGAGGCTGGGGCTGGAGCTGAGGCTGGGGCTGG 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1221 AGGTTGAGCTGGCGCTGTGTCTGTGGGGAATCTTCTGGGTCAAGGAGG 1162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 GGGCTGGGACTGAGGCTGGGGCTTANGGCTGGGGC 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1161 GGGCTGGGCTGAGGCTGAGGCTGATGCTGGGCC 1128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-08-655-241-17/c
; Sequence 17, Application us/08655241
; Patent No. 6025543
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
```

APPLICANT: Weigel, Detlef
TITLE OF INVENTION: Seed Plants Exhibiting Early Reproductive
TITLE OF INVENTION: Development and Methods of Making Same
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,241
FILING DATE: 05-JUN-1996
CLASSIFICATION: CLASS 800
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UD 1894
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1656 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1651
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..1656
OTHER INFORMATION: /note= "domain = ecdysone receptor
OTHER INFORMATION: ligand binding domain."
US-08-655-241-17
Query Match 20.8%; Score 104.2; DB 3; Length 1656;
Best Local Similarity 67.8%; Pred. No. 8.2e-14;
Matches 145; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
OY 10 GGGCTGGGCTGGGCTGAGGCTGAGGCTGAGGCTGGGCTGGGCTGGGCTG 69
DB 1341 GGGCAGCGAGCGAGACGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTG 1282
OY 70 GGGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 129
DB 1281 CGTGTGAGCTGTGCTGTGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1222
OY 130 AGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 189
DB 1221 AGGTTGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1162
OY 190 GGGCTGAGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 223
DB 1161 GGGCTGGGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1128
RESULT 5
US-09-398-326-17/c
Sequence 17, Application US/09398326
Patent No. 6355863
GENERAL INFORMATION:
APPLICANT: Yanofsky, Martin F.
TITLE OF INVENTION: Seed Plants Exhibiting Inducible Early
TITLE OF INVENTION: Reproductive Development and Methods of Making Same
NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/398,326
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/659,188
FILING DATE: 05-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UD 3739
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1656 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1651
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..1656
OTHER INFORMATION: /note= "domain = ecdysone receptor
OTHER INFORMATION: ligand binding domain."
US-09-398-326-17
Query Match 20.8%; Score 104.2; DB 4; Length 1656;
Best Local Similarity 67.8%; Pred. No. 8.2e-14;
Matches 145; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
OY 10 GGGCTGGGCTGGGCTGAGGCTGAGGCTGAGGCTGGGCTGGGCTGGGCTG 69
DB 1341 GGGCAGCGAGCGAGACGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTG 1282
OY 70 GGGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 129
DB 1281 CGTGTGAGCTGTGCTGTGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1222
OY 130 AGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 189
DB 1221 AGGTTGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1162
OY 190 GGGCTGAGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 223
DB 1161 GGGCTGGGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTG 1128
RESULT 6
US-09-144-759-17/c
Sequence 17, Application US/09144759
Patent No. 6117639
GENERAL INFORMATION:
APPLICANT: Hoock, Thomas
APPLICANT: Kwon, Ursula
TITLE OF INVENTION: FUSION PROTEINS, DNA MOLECULES, VECTORS, AND HOST CELLS
TITLE OF INVENTION: USEFUL FOR MEASURING PROTEASE ACTIVITY

```

1 FILE REFERENCE: VPI/98-08
2 CURRENT APPLICATION NUMBER: US/09/144,759
3 CURRENT FILING DATE: 1998-08-31
4 NUMBER OF SEQ ID NOS: 24
5 SOFTWARE: PatentIn Ver. 2.0
6 SEQ ID NO 17
7 LENGTH: 2241
8 TYPE: DNA
9 ORGANISM: Artificial Sequence
10 FEATURE:
11 OTHER INFORMATION: Description of Artificial Sequence:man-made
12 OTHER INFORMATION: artificial sequence
13 US-09-144-759-17

```

Query Match	20.8%	Score 104.2	DB 3	Length 2241
Best Local Similarity	67.8%	Pred. No. 8.6e-14		
Matches 145; Conservative	0	Mismatches 69	Indels 0	Gaps 0

[illegible]

```

RESULT 7
US-09-144-759-19/C
: Sequence 19, Application US/09144759
: Patent No. 6117639
: GENERAL INFORMATION:
: APPLICANT: Hoock, Thomas
: APPLICANT: Germann, Ursula
: APPLICANT: Kwong, Ann
: TITLE OF INVENTION: FUSION PROTEINS, DNA MOLECULES, VECTORS, AND HOST CELLS
: TITLE OF INVENTION: USEFUL FOR MEASURING PROTEASE ACTIVITY
: FILE REFERENCE: VPI/98-08
: CURRENT APPLICATION NUMBER: US/09/144,759
: CURRENT FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
:
: LENGTH: 2295

```

Query Match	20.8%	Score 104.2;	DB 3;	Length 2295;
Best Local Similarity	67.8%;	Pred. No. 8.7e-14;		
Matches 145; Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0;

[illegible]

Dd	1863	AGGTTTACTACCGGGCTGTGCTGTCGTCGGAATCGTTTGCTGAGGAGGAGGGGTG	1804
Qy	190	GGGCTGGACTGAGCTGGGCGCTAANGGCTGGGGGC	223
Dd	1803	GAGCTGGGCTBAGGCTBAGGCTGATCCTGGGCC	1770

```

RESULT 8
US-09-144-759-21/c
Sequence 21, Application US/09144759
Patent No. 6117639
GENERAL INFORMATION:
APPLICANT: Hooek, Thomas
APPLICANT: Hermann, Ursula
APPLICANT: Kwong, Ann
TITLE OF INVENTION: FUNCTIONAL PROTEINS, DNA MOLECULES, VECTORS, AND HOST CELLS
TITLE OF INVENTION: USEFUL FOR MEASURING PROTEASE ACTIVITY
FILE REFERENCE: VPI/98-08
CURRENT APPLICATION NUMBER: US/09/144,759
CURRENT FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO. 21
LENGTH: 2301
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:man-made
US-09-144-759-21

```

Query Match	20.8%	Score 104.2	DB 3	Length 2301
Best Local Similarity	67.8%	Pred. No. 8.7e-14		
Matches 145; Conservative	0	Mismatches 69	Indels 0	Gaps 0

[illegible]

```

1  RESULT 9
2  US-08-209-747-1
3  ; Sequence 1, Application US/08209747
4  ; Patent No. 5733771
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Lewis, Randolph V.
9  ;
10 ; APPLICANT: Colgin, Mark
11 ;
12 ; TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider
13 ;
14 ; NUMBER OF SEQUENCES: 56
15 ;
16 ; CORRESPONDENCE ADDRESS:
17 ;
18 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch
19 ;
20 ; STREET: P O Box 747
21 ;
22 ; CITY: Falls Church
23 ;
24 ; STATE: Virginia
25 ;
26 ; COUNTRY: USA
27 ;
28 ; ZIP: 22040-3487
29 ;
30 ; COMPUTER READABLE FORM:
31 ;
32 ; MEDIUM TYPE: Floppy disk
33 ;
34 ; COMPUTER: IBM PC compatible
35 ;
36 ; OPERATING SYSTEM: PC-DOS/MS-DOS

```


Db 904 CTGAGCCGCTGAGCTGAGAGAGAGAGAGCT 942

RESULT 11

PCT-US92-06391-1/c
 / Sequence 1, Application PC/TUS9206391
 / GENERAL INFORMATION:
 / APPLICANT: Genentech, Inc.
 / TITLE OF INVENTION: Ecdysteroid Dependent Regulation of
 / TITLE OF INVENTION: Genes In Mammalian Cells
 / NUMBER OF SEQUENCES: 1
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Genentech, Inc.
 / STREET: 460 Point San Bruno Blvd
 / CITY: South San Francisco
 / STATE: California
 / COUNTRY: USA
 / ZIP: 94080
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patin (Genentech)
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: PCT/US92/06391
 / FILING DATE: 19920803
 / CLASSIFICATION: 435
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Filtz, Renee A.
 / REGISTRATION NUMBER: 35,136
 / REFERENCE/DOCKET NUMBER: 607
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 415/225-1489
 / TELEFAX: 415/952-9881
 / TELETYPE: 910/371-7168
 / INFORMATION FOR SEQ ID NO: 1:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 2970 bases
 / TYPE: NUCLEIC ACID
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / PCT-US92-06391-1

Query Match 20.5%; Score 102.6; DB 5; Length 2970;
 Best Local Similarity 67.3%; Pred. No. 1.9e-13;
 Matches 144; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 10 GGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 69
 DB 2398 GGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 2339
 QY 70 GGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 129
 DB 2338 CCGCTGAGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 2279
 QY 130 AGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 189
 DB 2278 AGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 2219
 QY 190 GGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 223
 DB 2218 GGGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 2185

RESULT 12
 US-08-998-416-915/c
 / Sequence 915, Application US/08998416
 / Patent No. 6239264
 / GENERAL INFORMATION:
 / APPLICANT: Philippsen, Peter
 / APPLICANT: Rohlmann, Rainer
 / APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jurgen
 APPLICANT: Knechtel, Philipp
 APPLICANT: Redischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 915:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 688 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1570RP
 US-08-998-416-915

Query Match 20.4%; Score 101.8; DB 4; Length 688;
 Best Local Similarity 51.7%; Pred. No. 2.2e-13;
 Matches 211; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 60
 DB 609 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 550
 QY 61 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 120
 DB 549 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 490
 QY 121 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 180
 DB 489 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 430
 QY 181 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 240
 DB 429 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 370
 QY 241 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 300
 DB 369 CAGTTGTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 310
 QY 301 CTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 360
 DB 309 TTGAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 250
 QY 361 TGAAGCTGGGCTGGGCTGAGCTGAGCTGAGCTGAGCTGGGCTGGGCTGGGCTG 408

Fri Oct 11 09:30:52 2002

us-09-489-101a-17_copy_1_500.rni

Page 8

QY 315 GANGCTG 321
| |
| |
Db 304 GATTTTG 310

Search completed: October 10, 2002, 20:05:12
Job time : 24.7333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds
(without alignments)
9699.805 Million cell updates/sec

Title: US-09-489-101a-17_COPY_1_500
Perfect score: 500
Sequence: 1 ctgagcgcggcggcggcgtc.....gaatnaaatnccaccann 500

Scoring table: IDENTITY_NMC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	250.8	50.2	616	12	CNS03VHV
C 2	250	50.0	723	12	CNS03VHV
C 3	247.2	49.4	723	12	CNS02D1N
C 4	237.4	47.5	1061	12	CNS03Y86
C 5	237	47.4	880	12	CNS03Y86
C 6	237	47.4	949	12	CNS03Y86
C 7	226.2	45.2	971	12	CNS04AR5
C 8	225	45.0	966	12	CNS02SPN
C 9	224.6	44.9	890	12	CNS03SKB
C 10	221.8	44.4	965	12	CNS03E2K
C 11	221.6	44.3	899	12	CNS02E22
C 12	219.4	43.9	986	12	CNS01YMT
C 13	217.8	43.6	791	12	CNS02B0T
C 14	217.8	42.4	1056	12	CNS02B0T
C 15	210.2	42.0	1101	12	CNS04L67
C 16	194.6	38.9	906	12	CNS01S6G
C 17	193.4	38.7	475	12	CNS04S6N

C 18	187.8	37.6	721	12	CNS04208
C 19	184.8	37.0	975	12	CNS03750
C 20	183.8	36.8	959	12	CNS022KT
C 21	183.6	36.7	438	12	AZ858646
C 22	182.6	36.5	805	12	CNS04RW2
C 23	182.4	36.5	562	12	AZ855343
C 24	181.6	32.9	812	12	CNS04Q30
C 25	181.4	32.5	485	12	AZ434703
C 26	181.4	32.5	654	12	AZ937136
C 27	181.6	32.5	1091	12	CNS050EX
C 28	181.6	32.3	614	12	CNS01WDO
C 29	160	32.0	455	12	AQ928308
C 30	160	32.0	832	12	BG481596
C 31	159.8	32.0	1019	12	CNS02HKM
C 32	159	31.8	486	12	CNS01TND
C 33	158.6	31.7	865	12	CNS03ANA
C 34	157.6	31.5	832	12	CNS04FDB
C 35	152.6	30.5	485	9	AA930444
C 36	152	30.4	717	12	BH044236
C 37	150.4	29.9	655	9	BB628062
C 38	149.4	29.9	526	12	CNS03CMO
C 39	148.2	29.8	367	10	BF933221
C 40	148.6	29.8	533	12	AZ958534
C 41	148.6	29.7	961	12	CNS03OKT
C 42	147.6	29.5	837	10	BF697250
C 43	146.6	29.3	335	12	AZ415847
C 44	146.4	29.3	826	12	BH316049
C 45	146	29.2	372	10	BF932849

ALIGNMENTS

RESULT 1
LOCUS CNS03VHV/C
DEFINITION Tetraodon nigroviridis genome survey sequence pUC-ori end of clone 062P02 of library G from tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL262934.1 GI:7984575
VERSION AL262934.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 616) Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 616)
Bernot, A., Fizames, C., Winkler, P., Brothier, P., Quetier, F., Saurin, W. and Weissbach, J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 616)
REFERENCE
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
1..616
/organism="Tetraodon nigroviridis"

```

/db_xref="taxon:99883"
/clone="062P02"
/clone_11b="c"
/Note="Genoscope sequence ID : COB6062DH01SP1-end :
PUC-ori"

BASE COUNT      83 a      318 c      125 g      86 t      4 others
ORIGIN

Query Match      50.2%; Score 250.8; DB 12; Length 616;
Best Local Similarity 75.7%; Pred. No. 2.5e-29;
Matches 311; Conservative 1; Mismatches 98; Indels 1; Gaps 1;

OY 1 CTGAGGCTGGGGCTGGGGCTGGGGCTGAGCTGAGCTGAGCTGAGGCTGGGGCTGGGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 583 CTGAGCTCCGAGCCCGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 61 CTGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 523 CTGAGGCTGAGGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 121 CTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 463 CTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 181 CTGAGGCTGGGCTGGGCTGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCTAGG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 403 CCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 241 CTNAGGCTGAGGCTGGGGCTGGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 343 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 301 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 283 CTGAGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 360 CTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 223 CTGAGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGAGG 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||

RESULT 2
LOCUS      CENS01U74      723 bp      DNA      linear      GSS 12-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone
              197G24 of library G from Tetradon nigroviridis, genomic survey
              sequence.
ACCESSION   AL168241.1 GI:7806298
VERSION     AL168241.1
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetradon nigroviridis.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
              Tetraodontidae; Tetradon.
REFERENCE   1 (bases 1 to 723)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
              Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE       Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetradon nigroviridis
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 723)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A., Fizames,C., Mincker,P., Brottier,P., Quetier,F.,
              Saurin,W. and Weissenbach,J.
TITLE       Human gene number estimate provided by genome wide analysis using
              Tetradon nigroviridis DNA sequence
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 723)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A. and Weissenbach,J.
TITLE       Direct Submission

```

```

JOURNAL      Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT      This sequence is a single read and was generated as part of a large
              scale clone-end sequencing project of the Tetradon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/tetradon.
FEATURES
    source
    1..723
    /organism="Tetradon nigroviridis"
    /db_xref="taxon:99883"
    /clone="197G24"
    /clone_11b="c"
    /Note="Genoscope sequence ID : C0A6197BD12SP1-end :
    PUC-ori"

BASE COUNT      109 a      330 c      170 g      110 t      4 others
ORIGIN

Query Match      50.0%; Score 250; DB 12; Length 723;
Best Local Similarity 79.2%; Pred. No. 3.3e-29;
Matches 300; Conservative 1; Mismatches 77; Indels 1; Gaps 1;

OY 1 CTGAGGCTGGGGCTGGGGCTGGGGCTGAGGCTGGGACTGGGACTGGGCTGGGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 653 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 61 CTGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 593 CTGAGCTCCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 121 CTGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 533 CCGGGCTAGGCCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 181 CTGAGGCTGGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 473 CCGGGCTAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 241 CTNAGGCTAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 413 CCGAGGCTAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 301 CTGAGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 353 CTGAGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
OY 360 CTGAGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 293 CTGAGGCTGAGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||

RESULT 3
LOCUS      CENS02D1N      613 bp      DNA      linear      GSS 13-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone
              256B21 of library G from Tetradon nigroviridis, genomic survey
              sequence.
ACCESSION   AL191876.1 GI:7829980
VERSION     AL191876.1
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetradon nigroviridis.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
              Tetraodontidae; Tetradon.
REFERENCE   1 (bases 1 to 613)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
              Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE       Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetradon nigroviridis
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 613)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A. and Weissenbach,J.
TITLE       Direct Submission

```


[illegible]

LOCUS	GENE0318R	949 bp	DNA	linear	SS 15-MAY-2005
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 205022 of library G from Tetraodon nigroviridis, genomic survey sequence.				
ACCESSION	AL223901 GI:7882739				
VERSION	GSS; genome survey sequence.				
KEYWORDS	Tetraodon nigroviridis.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Tetraodoniformes; Neoteleostei; Tetraodonidae; Tetraodon.				
REFERENCE	1 (bases 1 to 949) Roest-Crolius H., Jallion O., Dasilva C., Fizesmes C., Fisher C., Bonneau L., Billault A., Queller F., Saurin W., Bernot A. and Weissenbach J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 949)				
AUTHORS	Roest-Crolius H., Jallion O., Dasilva C., Bonneau L., Fisher C., Bernot A., Fizesmes C., Winkler P., Brotlier P., Queller F., Saurin W. and Weissenbach J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 949)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases				
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.location/Qualifiers				
FEATURES	source				
BASE COUNT	205 a 202 c 390 g 144 t 8 others				
ORIGIN	Query Match 47.48; Score 237; DB 12; Length 949; Best local similarity 73.88; Pred. No. 3, 1e-27; Mismatches 99; Indels 1; Gaps 1.				
1	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
2	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
3	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
4	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
5	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
6	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
7	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
8	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
9	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
10	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
11	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
12	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
13	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
14	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
15	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
16	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
17	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
18	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
19	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
20	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
21	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
22	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
23	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
24	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				
25	CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCGTGGAGACTAGAGCTGGGGCTGGGG 60				

Oy 360 CTGAAGCTGAGGCTGGGGCTAACCTGAGCTGCGGCTGG 4000
 ||||| ||||| ||| ||
 Db 900 CTGAAGCTGAGGCGGAGGCTGAGCTGAGGCTGAAGCCGCGG 940

CNS044R5/c				
LOCUS	CNS044R5	971 bp	DNA	linear GSS 18-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone			

ACCESSION	AL274442
VERSION	AL274442.1
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetrahodon nigroviridis.
ORGANISM	Tetrahodon nigroviridis

REFERENCE
1 (bases 1 to 971)
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
AUTHORS

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 971)
AUTHORS	Roest-Crollius/H., Jallion/O., Dasliya/C., Bouneau/L., Fisher/C.,

TITLE	Human gene number estimate provided by genome wide analysis using tetraodon nigroviridis DNA sequence
JOURNAL REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 971)
JOURNAL TITLE	Genoscope.
COMMENT	Direct Submission Submitted (12-APR-2000) to the EMBL/GenBank/DBDy databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .

FEATURES	
source	Location/Qualifiers
	1. 971
	/organism="tetraodon nigroviridis"
	/db.xref="taxon:99883"
	/clone="082623"
	/clone.lib="G"
	/note="Genoscope sequence ID: COBG082ND12SP1-end PUC-Or"
BASE COUNT	156 a 356 c 247 g 205 t 7 others
ORIGIN	

Query Match	45.28;	Score 226.2;	DB 12;	Length 971;
Best Local Similarity	74.08;	Pred. No. 1.3e-25;		
Matches 290; Conservative	1;	Mismatches 100;	Indels 1;	Gaps 1;

QY

1 CTGAGCGCTGGGGGCTGGGGGCTAGAGGCCTGGAGCCTAGAGGCCTGGGGGCTGGGG 60
:||||| ||||| ||| ||||| |||| | ||| ||| ||| |||

61 CTGGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGACTGAGGCTGGGG 120

Db 851 CCGGGGGCCGGGACCGAGGCTGAGTCCGAGGCTGAGGCTGAGTCCGAGGCGCGGGGCTGAGG 792

[illegible]

181 CTGAGGTTGGGGCTGGGACTGAGGCTGGGGCTAGGCGTGGGGCTTAAAGG 240

[illegible]

QY 241 CTNAGGCTAGGTTGGGGGCTGGGGCTGGNGCTAACGCTGGGGCTAGGGCTGGNGCTAGG 300D
|| ||||| || ||||| || | ||||| || |||||
Db 671 CTGAGGCTGAGTCCAGAGGCCGGGGCTAGAGCCGGACCGAGGCTGAGGCTGAGGCTGAGG 612D

301 CTGAGCTGGGGCTGANGCTGGGGCTGGGGCTGNNCTGA - NCTGGGGCTGAGGCTTCNG 359

QY	360	CTGAAGCTGAGGCTGGGGC	NTAAGCTGAGCT	391
Db	551	CTGAGTCCGAGGCTGAGGCTGAGGT	CGGGGCT	520

CNS02SPN				
LOCUS	CNS02SPN	966 bp	DNA	linear
DEFINITION	Tetradon nigroviridis genome survey sequence T7 end of clone			

ACCESSION AL212180
VERSION AL212180.1 GI:7870999
KEYWORDS GSS, genome, survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis

REFERENCE
1 (bases 1 to 966)
Roest-Crollius, H., Jallou, O., Dasilva, C., Fizames, C., Fisher, C.,
Muller, J., and Oudiz, J. 1987. The structure of the 3' end of the
mRNA of the *Caenorhabditis elegans* *unc-54* gene. *Nucleic Acids Res.* 15:
1011-1021.

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Unpublished	2 (bases 1 to 966)	Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Flizanes, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.	Human gene number estimate provided by genome wide analysis using <i>Tetraodon nigroviridis</i> DNA sequence	Unpublished	3 (bases 1 to 966)

TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigriviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES	source	location/Qualifiers
1..966		1..966
		/organism="Tetrahodon nigroviridis"
		/db_xref="taxon:99883"
		/clone="162H05"
		/clone_1lb="G"
		/note="Genoscope sequence ID : COAG162C003LPI-end : T7"
BASE COUNT	206 a	245 c 359 g 153 t 3 others
ORIGIN		

Query Match	45.0%;	Score 225;	DB 12;	Length 966;
Best Local Similarity	74.0%;	Pred. No. 2e-25;		

1 CTGAGGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGG 60

Db 132 CTGAGGCTGAGTCCGAGGCCCGGGGCTGAGGCTGAGTCCGAGTCCGAGGCCCGGGGCTGAGG 19

Dbb 192 CTGAGCGCGGGCTAGGCTCCGAGCCCGGGCAGAGGCCGGGACCGAGGCTAGG 25

QY 121 CTGGGCGTGAAGCTGGGGCTGGGACTGAGGCTGGGCTGGGGCTGGGG 18

Db 252 CTGAGTCCGAGGCGGGGCTGAGCTGAGCCAGGCGGGGCTGAGGCTGAGGCTGAGG 311
 181 CTGAGCTGGGGCTGAGGACTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 240
 Db 312 CTGAGCTCCGAGGCGGGGCTGAGGCGGGGCTGAGGCGGGGCTGAGGCTGAGG 371
 Oy 241 CTNAGGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 300
 Db 372 CTGAGCTGAGGCGGGGCTGAGGCGGGGCTGAGGCTGAGGCTGAGGCTGAGG 431
 Oy 301 CTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 359
 Db 432 CTGAGGCGGGGCTGAGGCTGAGGCGGGGCTGAGGCTGAGGCTGAGGCTGAGG 491
 Oy 360 CTGAGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 391
 Db 492 CCGGGGCTGAGGCGGGGCTGAGGCGGGGCTGAGGCGGGGCTGAGGCGGGGCT 523
 RESULT 9
 CNS035KB/c
 LOCUS
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Orl end of clone
 214609 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 890 bp DNA linear GSS 15-MAY-2000
 ACCESSION AL228836.1 GI:7887828
 VERSION AL228836.1
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 890)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetler,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 890)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brotlier,P., Quetler,F.,
 Saurin,W. and Weissenbach,J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 890)
 AUTHORS Direct Submission
 TITLE Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 1. 890
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone.lib="G"
 /note="Genoscope sequence ID : C0AG214AD05SP1-end :
 PUC-Orl"
 BASE COUNT 139 a 368 c 203 g 154 t 26 others
 ORIGIN
 Query Match 44.9%; Score 224.6; DB 12; Length 890;
 Best local Similarity 70.9%; Pred. No. 2.4e-25;
 Matches 292; Conservative 5; Mismatches 114; Indels 1; Gaps 1;
 Oy 1 CTNAGGCTGGGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGGGG 60

Db 885 CTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 826
 Oy 61 CTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 120
 Db 825 CTGAGCTCCGAGGCGGGGCTGAGGCGGGGCTGAGGCGGGGCTGAGGCTGAGG 766
 Oy 121 CTGAGGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 180
 Db 765 TCGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 706
 Oy 181 CTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 240
 Db 705 CCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 646
 Oy 241 CTNAGGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 300
 Db 645 CCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 586
 Oy 301 CTGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 359
 Db 585 CTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 526
 Oy 360 CTGAGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 411
 Db 525 CTGAGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 474
 RESULT 10
 CNS012DK
 LOCUS
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 220005 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 965 bp DNA linear GSS 12-MAY-2000
 ACCESSION AL174161.1 GI:7812218
 VERSION AL174161.1
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 965)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetler,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 965)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brotlier,P., Quetler,F.,
 Saurin,W. and Weissenbach,J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 965)
 AUTHORS Direct Submission
 TITLE Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 1. 965
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone.lib="G"
 /note="Genoscope sequence ID : C0AG220AH03JP1-end : T7"
 BASE COUNT 170 a 335 c 224 g 232 t 4 others

JOURNAL REFERENCE AUTHORS TITLE	JOURNAL TITLE AUTHORS COMMENT	FEATURES source	BASE COUNT ORIGIN
freshwater pufferfish Tetraodon nigroviridis Unpublished 2 (bases 1 to 1056) Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysts using Tetraodon nigroviridis DNA sequence Unpublished 3 (bases 1 to 1056) Genoscope. Direct Submission Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon . location/Qualifiers 1. 1056 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="167P12" /clone_1b="G" /note="Genoscope sequence ID : COAG167DC06P1-end : T7" <td>190 a 358 c 273 g 234 t 1 others</td>	190 a 358 c 273 g 234 t 1 others		
Query Match Best Local Similarity Matches 278; Conservative	42.4%; Score 211.8; DB 12; Length 1056; 72.8%; Pred. No. 2.1e-23; 0; Mismatches 103; Indels 1; Gaps 1;		
OY 1 CTGAGGCTGGGGCTGGGGCTGGGGCTGAGCTGAGCTGAGGCTGGGGCTGGGG 60 DB 1006 CCGGGCGAGAGCCGGGGCTAGGCTAGGCTGGGGCTGAGGCTGAGGCTGAGT 947 OY 61 CTGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGAGCTGAGG 120 DB 946 CCGAGGCCGGGGCCGAGGCTGAGTCCGAGGCCGGGGCTGAGGCTGAGGCTGAGT 887 OY 121 CTGGGGCTGAGGCTGGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGAGGCTGGGG 180 DB 886 CCGAGGCCGGGGCCGAGGCTGAGTCCGAGGCCGGGGCTGAGGCTGAGGCTGAGT 827 OY 181 CTGAGCTGGGGCTGGGACTGAGGCTGGGGCTGAGGCTGGGGCTGAGGCTGAGG 240 DB 826 CTGAGGCTGAGTCCGAGGCCGGGGCCGAGGCTGAGTCCGAGGCTGAGGCTGAGG 767 OY 241 CTNAGGCTGAGGTTGGGGCTGGGGCTGGGGCTGAGGCTGGGGCTGAGGCTGGGG 300 DB 766 CTGAGGCTGAGTCCGAGGCTGAGGCTGAGTCCGAGGCCGGGGCTGAGGCTGAGG 707 OY 301 CTGAGGCTGGGGCTGAGGCTGGGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGG 359 DB 706 CTGAGGCTGAGGCTGAGGCTGGGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGG 647 OY 360 CTGAGCTGAGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 381 DB 646 CCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 625			
RESULT 15 CONS04LE7/c LOCUS DEFINITION	1101 bp DNA linear GSS 21-MAY-2000 Tetraodon nigroviridis genome survey sequence PUC-Orl end of clone 118102 of library G from Tetraodon nigroviridis, genomic survey sequence.		
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AL296008.1 GI:8034588 GSS: genome survey sequence. Tetraodon nigroviridis. Tetraodon nigroviridis Tetraodon nigroviridis Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 84.9333 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-17_COPY_1_500

Sequence: 1 ctagggctggggctggggct.....gaatnaaatccacccam 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Genesec_032802: *
2: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1980.DAT: *
3: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1981.DAT: *
4: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1982.DAT: *
5: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1983.DAT: *
6: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1984.DAT: *
7: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1985.DAT: *
8: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1986.DAT: *
9: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1987.DAT: *
10: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1988.DAT: *
11: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1989.DAT: *
12: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1990.DAT: *
13: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1991.DAT: *
14: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1992.DAT: *
15: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1993.DAT: *
16: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1994.DAT: *
17: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1995.DAT: *
18: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1996.DAT: *
19: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1997.DAT: *
20: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1998.DAT: *
21: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA1999.DAT: *
22: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA2000.DAT: *
23: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA2001A.DAT: *
24: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA2001B.DAT: *
25: /SIDSI/gcgdata/hold-genesec/genesecq-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	460	92.0	1232	22	AA011124
2	171	34.2	2465	11	AA003665
3	165.8	33.2	564	22	ABA00214
4	165.8	33.2	564	22	ABA28356
5	165.8	33.2	564	22	AAK08552
6	165.8	33.2	564	22	AAK34435
7	165.8	33.2	564	22	AAI16734
8	165.8	33.2	564	22	AAI40161
9	163.4	32.7	569	23	AA584303

10	163.4	32.7	1209	23	AA580389
11	163.4	32.7	1209	23	AA581401
12	161.6	32.3	10347	22	AA280600
13	148.4	29.7	368	22	ABA72613
14	148.4	29.7	368	22	ABA58433
15	148.4	29.7	368	22	AAK21244
16	148.4	29.7	368	22	AAK47401
17	148.4	29.7	368	22	AAI25929
18	148.4	29.7	368	22	AAI53240
19	147	29.4	2403	23	ABL24558
20	144.2	28.8	3092	23	AA569141
21	144.2	28.8	10578	23	AA584306
22	143.2	28.6	1548	12	AAQ13868
23	143	28.6	303	23	ABL24559
24	142.8	28.6	3526	12	AAQ13864
25	137.8	27.6	1172	24	AA023835
26	135.8	27.2	716	21	AA44382
27	132.4	26.5	521	21	AA066020
28	132.4	26.5	2708	22	AAH34279
29	131	26.1	2248	24	AA023837
30	130.4	26.1	1954	22	AAI59802
31	129.2	25.8	486	22	AAI75507
32	127.8	25.6	5120	22	AA084677
33	127.4	25.5	6794	21	AA245597
34	125.6	25.1	3780	24	AA520005
35	122.8	24.6	4279	23	AA586600
36	121.4	24.3	2550	19	AAV52497
37	120.8	24.2	6558	21	AA245602
38	119.6	23.9	590	24	AA023836
39	115.4	23.1	2465	11	AA003665
40	114.8	22.9	2668	22	AAK75409
41	114.6	22.9	3975	16	AA092412
42	113.8	22.8	383	16	AA088689
43	113.8	22.8	666	16	AA088690
44	113.8	22.8	807	19	AAV52505
45	113.8	22.8	3633	21	AA05548

ALIGNMENTS

RESULT 1
AAD11124 standard; DNA: 1232 BP.
XX
AC AAD11124:
XX
DT 24-SEP-2001 (first entry)
XX
DE Human small cell lung cancer associated gene, Novel-2.
XX
KM Human: small cell lung cancer; therapy: hCAAP; nucleic acid: NA;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW leiomyosarcoma; synovial sarcoma; cytoskeletal; ds.
XX
OS Homo sapiens.
XX
PN WO200153349-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001MO-US02015.
XX
PR 21-JAN-2000; 2000US-0489101.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (SICK) SLOAN KETTERING INST CANCER RES.
PA (CORR) CORNELL RES FOUND INC.
XX
PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX
DR WPI: 2001-457597/49.
XX

PT Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e.9. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 PS Claim 57; Page 144-145; 152pp; English.
 XX
 CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is human Novel-2 DNA, a small cell lung cancer
 CC associated gene designated as NY-SCLC-14.
 XX
 SQ Sequence 1332 BP; 249 A; 212 C; 261 G; 243 T; 267 other;
 Query Match 92.0%; Score 460; DB 22; Length 1232;
 Best Local Similarity 100.0%; Pred. No. 6,9e-72;
 Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGAGGCTGGGGCTGGGGCTGGAGCTGAGCTGGAGCTGAGGCTGGGGCTGGGG 60
 DB 1 CTGAGGCTGGGGCTGGGGCTGGAGCTGAGCTGGAGCTGAGGCTGGGGCTGGGG 60
 QY 61 CTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
 DB 61 CTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
 QY 121 CTGGGGCTGGGGCTGGGGCTGGAGCTGAGGCTGGAGCTGAGGCTGGGGCTGGGG 180
 DB 121 CTGGGGCTGGGGCTGGGGCTGGAGCTGAGGCTGGAGCTGAGGCTGGGGCTGGGG 180
 QY 181 CTGAGCTGGGGCTGGAGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 240
 DB 181 CTGAGCTGGGGCTGGAGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 240
 QY 241 CTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 300
 DB 241 CTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 300
 QY 301 CTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 360
 DB 301 CTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 360
 QY 361 TGAAGCTGAGGCTGGGGCTGAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGG 420
 DB 361 TGAAGCTGAGGCTGGGGCTGAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGG 420
 QY 421 TGAAGCTGAGGCTGGGGCTGAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGG 480
 DB 421 TGAAGCTGAGGCTGGGGCTGAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGG 480
 QY 481 GAATNNAATNTCCACCA 498
 DB 481 GAATNNAATNTCCACCA 498
 RESULT 2
 ID AA003665/C
 AA003665 standard; DNA; 2465 BP.
 XX
 AC AA003665;
 XX
 DT 07-AUG-1989 (first entry)
 XX
 DE Sequence homologous to Drosophila Per gene.
 XX

KW Sex determination; ruminant embryos; ss.
 XX
 PN FR2635116-A.
 XX
 PD 09-FEB-1990.
 XX
 PF 08-AUG-1988; 88FR-0010706.
 XX
 PR 08-AUG-1988; 88FR-0010706.
 XX
 PA (GEOR/) GEORGES M.
 XX
 PI Georges M, Vassart G, Christophe D, Dumont J, Young M;
 XX WPI, 1990-093373/13.
 XX
 DR WPI, 1990-093373/13.
 XX
 PT Determ. of sex of ruminant embryos by hybridisation assay using specific
 PT probes.
 PS
 PS Claim 17; fig 1; 37pp; French.
 CC
 CC This sequence is used as a probe and is contained in plasmid pSP64.2 SEI.
 CC It is homologous to the sequence of a 2.5 kb fragment of the Per gene of
 CC Drosophila. The motifs ACGGN and/or TCAAGC (M-T, U, G, A, or C) or their
 CC complementary sequences are repeated in the probe. It is used in a
 CC hybridisation assay to determine the sex of ruminant embryos. Either
 CC (i) Southern- or (ii) Dot-blot techniques can be utilised and X-chromo-
 CC some shows up as intense signal either obscuring most of the hybridisa-
 CC tion track in (i) or surrounding dots in (ii). Significant results can
 CC be obt. using samples of less than 50 ng weight.
 CC
 SQ Sequence 2465 BP; 791 A; 739 C; 689 G; 246 T; 0 other;
 Query Match 34.2%; Score 171; DB 11; Length 2465;
 Best Local Similarity 62.2%; Pred. No. 1.2e-21;
 Matches 253; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
 QY 1 CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCTGAGCTGGAGCTGGGGCTGGGG 60
 DB 2061 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGAGCTGAGCTGGAGCTGGGGCTGGGG 2002
 QY 61 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
 DB 2001 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1942
 QY 121 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 180
 DB 121 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 180
 QY 1941 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1882
 DB 1941 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1882
 QY 181 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 240
 DB 181 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 240
 QY 1881 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1822
 DB 1881 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1822
 QY 241 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 300
 DB 241 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 300
 QY 1821 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1762
 DB 1821 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1762
 QY 301 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 360
 DB 301 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 360
 QY 1761 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1702
 DB 1761 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1702
 QY 361 TGAAGCTGAGGCTGGGGCTGAACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGG 410
 DB 1701 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1652
 RESULT 3
 ID ABA60274
 ABA60274 standard; DNA; 564 BP.
 XX
 AC ABA60274;
 XX
 DT 01-FEB-2002 (first entry)
 XX

XX DE Human foetal liver single exon nucleic acid probe #8579.
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX OS Homo sapiens.
 XX PN WO200157277-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00669.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-483447/52.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human fetal liver -
 XX PS Claim 1; SEQ ID NO 8579; 639pp + sequence listing; English.
 XX SS
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC XX
 SQ Sequence 564 BP; 75 A; 103 C; 227 G; 159 T; 0 other;
 Query Match 33.2%; Score 165.8; DB 22; Length 564;
 Best Local Similarity 64.7%; Pred. No. 9.5e-21;
 Matches 264; Conservative 0; Mismatches 142; Indels 2; Gaps 2;
 QY 1 CTGAGGCTGGGGGCTGGGGGCTGGAGCTGGAGCTGGAGCTGGGGGCTGGGG 60
 DB 16 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGCC 75
 QY 61 CTGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 120
 DB 76 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGCC 134
 QY 121 CTGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 180
 DB 135 CTGGGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGG 194
 QY 181 CTGAGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTAGGG 240
 DB 195 CTATGAGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGG 253
 QY 241 CTGAGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 300
 DB 254 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGCC 313
 QY 301 CTGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 360
 DB 314 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGCC 373
 QY 361 TGAAGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 408

DB 374 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGG 421
 RESULT 4
 ID ABA28556
 AC ABA28556 standard; DNA; 564 BP.
 XX
 AC ABA28556;
 XX
 PT 23-JAN-2002 (first entry)
 XX
 DE Probe #7022 for gene expression analysis in human heart cell sample.
 XX
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX
 OS Homo sapiens.
 XX PN WO200157274-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00666.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488899/53.
 XX PT Single exon nucleic acid probes for analyzing gene expression in human
 XX PT hearts -
 XX PS Claim 1; SEQ ID NO 7022; 530pp; English.
 XX SS
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC XX
 SQ Sequence 564 BP; 75 A; 103 C; 227 G; 159 T; 0 other;
 Query Match 33.2%; Score 165.8; DB 22; Length 564;
 Best Local Similarity 64.7%; Pred. No. 9.5e-21;
 Matches 264; Conservative 0; Mismatches 142; Indels 2; Gaps 2;
 QY 1 CTGAGCTGGGGGCTGGGGGCTGGAGCTGGAGCTGGAGCTGGGGGCTGGGG 60
 DB 16 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGCC 75
 QY 61 CTGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 120
 DB 76 CTGTGGCTATGAGCTGGGACCTGTGAGCTGTGGCTGGGACCTGTGAGCTGGCC 134
 QY 121 CTGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGGCTGGGGG 180

Db 5506 GAGGTGAAAGTGTGAGTGAAGCC 5483

RESULT 13
ID ABA72813 standard; DNA: 368 BP.

AC ABA72813;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #21118.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN W0200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 21118; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 368 BP; 29 A; 76 C; 149 G; 114 T; 0 other;

XX Query Match 29.7%; Score 148.4; DB 22; Length 368;

XX Best Local Similarity 66.7%; Pred. No. 1e-17;

XX Matches 234; Conservative 0; Mismatches 115; Indels 2; Gaps 2;

XX

XX

XX

XX

OY 241 CTNAGCTGAGTGTGGGCTGGGCTGACGCTGAGGCTGAGGCTGAGGCTGAGG 300

DB 255 CTGAGCTGTGCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 314

OY 301 CTGAGCTGTGCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 351

DB 315 CTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 365

RESULT 14

ID ABA38433 standard; DNA: 368 BP.

AC ABA38433;

DT 23-JAN-2002 (first entry)

DE Probe #16899 for gene expression analysis in human heart cell sample.

KW Human; gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease; ss.

OS Homo sapiens.

PN W0200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

XX Claim 4; SEQ ID NO 16899; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart. The

XX present sequence is one such probe. The probes may be used for

XX predicting, measuring and displaying gene expression in samples derived

XX from the human heart via microarrays. By measuring gene expression, the

XX probes are useful for predicting, diagnosing, grading, staging,

XX monitoring and prognosing diseases of the human heart and vascular system

XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

XX congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 368 BP; 29 A; 76 C; 149 G; 114 T; 0 other;

XX Query Match 29.7%; Score 148.4; DB 22; Length 368;

XX Best Local Similarity 66.7%; Pred. No. 1e-17;

XX Matches 234; Conservative 0; Mismatches 115; Indels 2; Gaps 2;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Fri Oct 11 09:30:51 2002

us-09-489-101a-17_copy_1_500.rng

Page 10

```

OY 61 CTGGGCTGGGCTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGAGCTGGGG 120
Db 76 CTGGTGTATGGCTGTAGCTGGGACCTGGCTGTGGCTGTGAGTGGCACTGTGGCTGTGG 135
OY 121 CTGGGCTGAGGCTGGGGCTGGGACTGAGCTGGGGCTGGGACTGAGGCTGGGGCTGGGG 180
Db 136 CTATGACTGTACTGTGCTGTGG-CTGGTCTATGGCTGTAGCTGGACCTGGCTGTGG 194
OY 181 CTGAGTTGGGGCTGGGACTGAGGCTGGGGCTGANGGCTGGGGCTGAGGCTGGGCTAGGG 240
Db 195 CTGGGCTCTGACTGGCCCTGTGGCTGTGGCTGTGGCTGTGGATGTAGCTGGCACTGTGG 254
OY 241 CTNAGCTGAGGCTGGGGCTGGGGCTGAGCTGAGCTGGGGCTGAGGCTGGGAGCTAGG 300
Db 255 CTGAGCTGGTGTGCTGAGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGGCCCTGTGG 314
OY 301 CTGAGCTGGGGCTGANGCTGGGGCTGGGGCTGANNCTGANTCTGANGCTGGGCTGA 351
Db 315 CTGGGCTATGACTGGGACCTGGCCCTGGCTGTGGCTGTGGCTGTGGTGTGA 365

RESULT 15
AAK21244
ID AAK21244 standard; DNA: 368 BP.
XX
AC AAK21244;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 21235.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
XX MO200157275-A2.
PN
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI
PL Penn SG, Hanzel DK, Chen W, Rank DR:
PT WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 21235; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 368 BP; 29 A; 76 C; 149 G; 114 T; 0 other:
XX

```

Query Match	29.7%	Score 148.4	DB 22	Length 368
Best Local Similarity	66.7%	Posed No. 1e-17		
Matches 224	Conservative 0	Mismatches 115	Indels 2	Gaps 2
QY	1	CTGAGGCTGAGGCGCTGGGGCGTGAAGCGTGAAGCGTGGGCGTGGGCG	60	
Db	17	CTGTGGCTGTGGCTGTGAAGCTGTGGGCTGTGGGCTATGACGTAACTGTGTCTTGG	75	
QY	61	CTGGGGCTGGGGCGTGAAGCTGGGGGCGTGGGGCGTGGGCGCTGGGACTGAAGCGTGGGG	120	
Db	76	CTGGTCTATGAGGCTGTGATGCTGGCACTGTGGCTGTGGCTGTGAAGGCACTGTGGGCTGGG	135	
QY	121	CTGGGGCGTGAAGCGTGGGGCGTGGGCACTGAAGCGTGGGGCGTGGGACTGTGGGCGTGGGG	180	
Db	136	CTATGACTGTGTACTGTGTGTCTTGG-CTGGTGGCTAAGGCGCTGTAAGCTGGCACTGTGGGCTTGG	194	
QY	181	CTGAGGTTGGGGCTGGGAGCTGAAGGCTGGGGCTAAGCGTGGGCGCTGAAGCTGGGGCTAAGG	240	
Db	195	CTGTGGCTCTGTGACTGGGCCCTGTGGGCTGTGGGCGTGGGCTGTGGATGTAGCGCTGGCACTGGG	254	
QY	241	CTTAAAGCTAGAGTTGGGGCGTGGGGCTGGGAGCTGAAGCGTGGGGCGCTGAAGCTGAAGCTGAAG	300	
Db	255	CTGTACTGTGGCTGTGTGACTGTGGCTGTGGCTGGGCGTGGGCGTGTGACTGTGGCCCTGGG	314	
QY	301	CTTGAAGCTGGGGCGTGAAGCGTGGGGCGTGGGGCGTGGGCGTGAAGCTGAAGCTGAAGCTGAAG	351	
Db	315	CTGTGGCTATGACTGTGGCAAGCGTGGGCGTGGGCGTGGGCGTGGGCGTGGGCGTGGGCGTGGG	365	

Search completed: October 10, 2002, 15:18:11
Job time : 91 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds

(network configuration)
14616.225 million cell updates/sec

Title: US-09-489-101A-17_COPY_1_500

```

Sequence: 1 ctgagctgggctgggct.....gattnaaatntccaccann 500

```

Scoring table: IDENTITY_NUC

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  Sentence: *
2:      1:  gb_ba: *
3:      2:  gb_htg: *
4:      3:  gb_in: *
5:      4:  gb_om: *
6:      5:  gb_ov: *
7:      6:  gb_pat: *
8:      7:  gb_ph: *
9:      8:  gb_pl: *
10:     9:  gb_pr: *
11:    10:  gb_ro: *
12:    11:  gb_stg: *
13:    12:  gb_sy: *
14:    13:  gb_un: *
15:    14:  gb_vl: *
16:    15:  em_ba: *
17:    16:  em_fun: *
18:    17:  em_hum: *
19:    18:  em_in: *
20:    19:  em_mu: *
21:    20:  em_om: *
22:    21:  em_or: *
23:    22:  em_ov: *
24:    23:  em_pat: *
25:    24:  em_ph: *
26:    25:  em_pl: *
27:    26:  em_ro: *
28:    27:  em_stg: *
29:    28:  em_sy: *
30:    29:  em_un: *
31:    30:  em_htg_hum: *
32:    31:  em_htg_inv: *
33:    32:  em_htg_other: *
34:    33:  em_htg_gov_inv: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

1	460	92.0	1332	6	AX201596	Sequence		
C	2	289.4	57.9	54587	2	AC094895	Rattus no	
3	285	57.0	143720	10	AC089833	AC0987332	Mus muscu	
4	280.4	56.1	220761	2	AE670227	AE670227	Mus muscu	
5	258.4	51.7	2801	5	AF1281784	AF1281784	Gallus ga	
6	248.8	49.8	172224	2	AC0871149	AC0871149	Mus muscu	
C	7	225	45.0	118562	9	AC0266443	Homo sapi	
8	225	44.2	134909	2	AC025183	AC025183	Homo sapi	
9	220.8	44.2	179386	3	AC104515	AC104515	Drosophill	
10	219.4	43.9	182421	2	AC00811760	AC00811760	Drosophill	
11	217.6	43.5	179595	2	AC098110	AC098110	Rattus no	
C	12	215	43.0	32773	10	AF1189283	Mus muscu	
13	215	43.0	110312	10	AF1189286	AF1189283	Mus muscu	
C	14	214.8	43.0	7085	2	AC013260	Drosophill	
C	15	214.8	43.0	274388	3	AB003574	Drosophill	
16	213.8	42.8	73038	2	AC0208884	AC0208884	Drosophill	
C	17	213.4	42.7	1935	10	AF18926452	Continuation (4 of	
18	209.8	41.8	122592	3	CE943A2	AF189265	Mus muscu	
C	19	208.8	41.8	39498	2	AF189262	AF189262	Caenorhab
C	20	205.4	40.1	3000	10	AF189232	AF189262	Rattus no
C	21	205.4	40.1	290452	2	AC079167	AC079167	Mus muscu
22	204.2	40.8	139934	9	H0B384D8	H0B384D8	Mus sapien	
C	23	203.4	39.7	3978	10	AF255385	AF255385	Rattus no
24	199	39.8	110000	2	CEY102631	CEY102631	Continuation (2 of	
C	25	197.8	39.6	165849	2	AC096220	AC096220	Rattus no
C	26	195.2	39.0	192202	2	AC094105	AC094105	Homo sapi
27	195	39.0	167224	4	AC095838	AC095838	Rattus no	
28	194.4	38.9	67391	2	AC101718	AC101718	Mus muscu	
29	193.6	38.7	148871	2	AC025324	AC025324	Homo sapi	
C	30	193	38.6	3708	9	AC015261	AC015261	Drosophill
31	191.8	38.4	146339	9	H0MBAC	H0MBAC	L49235	
C	32	191.4	38.3	154825	2	AC1103031	AC1103031	Rattus no
33	189.6	37.7	34727	3	CE111A6	CE111A6	Caenorhabdi	
C	34	187	37.4	89370	3	AC084153	AC084153	Caenorhab
35	186.2	37.2	226502	2	AL645802	AL645802	Mus muscu	
36	186	37.2	1135	10	M0SHSV1B	M1043	Mouse DNA f	
C	37	185.4	37.1	257817	2	AC006909	AC006909	Caenorhab
C	38	184.4	36.9	66441	2	AC0065715	AC0065715	Caenorhab
39	183.6	36.7	558	3	TERSTELA	M1622	T. thermophil	
C	40	183.6	36.7	191692	3	AC020834	AC020834	Mus muscu
41	183	36.6	157909	2	AC096043	AC096043	Rattus no	
C	42	181.4	36.3	34277	3	CE11A16	CE11A16	Caenorhabdi
43	181.2	36.2	2679	3	CE11A33B	CE11A33B	Caenorhab	
C	44	179.6	35.9	37569	2	CE26H2	CE26H2	R28205
45	177.6	35.9	43255	2	CEC37B6	CEC37B6	Caenorhabdi	

[illegible]

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550). Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences. Genes and Regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST and cDNA sequences in Unigene. Genes demonstrate at least two exons flanked by consensus splice sites that maintain sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double stranded sequence for all regions. All sequence is completed to a standard of coverage with a minimum of 3 reads with no ambiguities. If the sequence coverage for a region does not meet this standard, it is indicated in the annotation as Low Coverage. Low coverage linkages are verified by PCR product size verification or verification of forward and reverse reads from clones which span the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards. Estimated average error rate is less than 1 per 10,000 bases using the Consed quality parameters. Regions that do not meet this requirement are annotated as Low Quality.

Summary Statistics

Center project name: AJP
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 1008
Assembly program: Phrap version 0.990319
Contig length: 143720
Fraction of Phrap value < 40: 8.349e-05
Error Rate in Consed: 0.00 per 10,000 bases
Number of N's in consensus: 0

Distribution of Quality < 40 Bases:

```

10001
9001
8001
7001
6001
5001
4001
3001
2001
1001
01

```

```

5 10 15 20 25 30 35 40
Phrap Value Range

```

FEATURES

```

source
-----
Location/Qualifiers
1. 143720
/organism="Mus musculus"
/db_xref="taxon:10090"
repeat_region
2373..2464
/rpt_family="L3"
complement(3181..3264)
repeat_region
3265..3380
/rpt_family="L1MB7"
complement(3381..3819)
repeat_region
3834..3914
/rpt_family="L1MB7"
complement(3946..4014)
repeat_region
/rpt_family="B4A"

```

```

repeat_region complement(4015..4232)
repeat_region /rpt_family="B3A"
repeat_region complement(4233..4449)
repeat_region /rpt_family="B4A"
repeat_region 4529..4567
repeat_region /rpt_family="Cn-rich"
repeat_region complement(4636..4779)
repeat_region /rpt_family="B1_MM"
repeat_region complement(4957..5065)
repeat_region /rpt_family="L1M4"
repeat_region 7137..7206
repeat_region /rpt_family="PB1D10"
repeat_region 7208..7231
repeat_region /rpt_family="(CANA)n"
repeat_region 7682..7906
repeat_region /rpt_family="(GCA)n"
repeat_region 8548..8632
repeat_region /rpt_family="B1_MM"
repeat_region 10714..10795
repeat_region /rpt_family="(CAGAGA)n"
repeat_region complement(10946..11091)
repeat_region /rpt_family="B1_MM"
repeat_region 11502..11580
repeat_region /rpt_family="(TG)n"
repeat_region 12288..12443
repeat_region /rpt_family="RSINE1"
repeat_region complement(12521..12648)
repeat_region /rpt_family="MIR"
repeat_region complement(15305..15523)
repeat_region /rpt_family="MIR"
repeat_region 17262..17351
repeat_region /standard_name="456819"
repeat_region complement(19924..20048)
repeat_region /rpt_family="B1F"
repeat_region 26227..26603
repeat_region /rpt_family="(CTGGGG)n"
repeat_region 27851..27887
repeat_region /rpt_family="(CA)n"
repeat_region 27887..27921
repeat_region /rpt_family="(GA)n"
repeat_region 29704..29733
repeat_region /rpt_family="(TA)n"
repeat_region 30264..30401
repeat_region /rpt_family="B1_MM"
repeat_region 33524..33686
repeat_region /rpt_family="B2_MM1"
repeat_region 33687..33709
repeat_region /rpt_family="(TAAA)n"
repeat_region 33878..34603
repeat_region /standard_name="D2Ar218"
repeat_region complement(35888..36002)
repeat_region /rpt_family="PB1D7"
repeat_region 36558..36822
repeat_region /rpt_family="L1M4"
repeat_region 36830..36938
repeat_region /rpt_family="PB1D7"
repeat_region 36962..37038
repeat_region /rpt_family="L1M4"
repeat_region 38396..38754
repeat_region /rpt_family="MTD"
repeat_region 39284..39405
repeat_region /rpt_family="PB1D10"
repeat_region 39406..39431
repeat_region /rpt_family="(GGA)n"
repeat_region 39448..39650
repeat_region /rpt_family="B3A"
repeat_region 39771..40266
repeat_region /rpt_family="L1M4"
repeat_region 40273..40321
repeat_region /rpt_family="(TG)n"
repeat_region complement(40705..40894)

```

Query Match

57.0%; Score 285; DB 10; Length 143720;

```

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: BM317N1
-----
Summary Statistics
Assembly program: XGAP4: version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 219697 bases at least Q40
Consensus quality: 219887 bases at least Q30
Consensus quality: 220081 bases at least Q20
Insert size: 220261; sum-of-contents
Insert size: 213159; 11.3% error; agarose-fp
Quality coverage: 13.01x in Q20 bases; sum-of-contents Quality
coverage: 13.51x in Q20 bases; agarose-fp

```

FEATURES	Location/Qualifiers
source	1. .220761 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="4" /clone="RP23-317N1" /clone_1lib="RPci-23" 1. .129995 /note="assembly:fragment:03451 fragment_chain:1" 130096..133380 /note="assembly:fragment:02708 fragment_chain:1" 135481..150170 /note="assembly:fragment:00061 fragment_chain:1" 150271..156339 /note="assembly:fragment:01326 fragment_chain:1" 156640..208913 /note="assembly:fragment:02733 fragment_chain:1" 209014..220761 /note="assembly:fragment:05258 fragment_chain:1 clone_end:sp6 vector_side:right"
BASE COUNT	57530 a 54793 c 52961 g 54975 t 502 others
ORIGIN	
Query Match	56.1%; Score 280.4; DB 2; Length 220761; Best local similarity 83.5%; Pred. No. 5.4e-39;
Matches 344; Conservative 0; Mismatches 65; Indels 3; Gaps 3;	
OY 1	CTGAGGCTGAGGCTGGGGCTGGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGGGG 60
Db 119954	CTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGGGG 119955
OY 61	CTGGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119
Db 119894	CTGGGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119895
OY 120	GTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 179
Db 119834	GTGAGGCTGGGGCTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119775
OY 180	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 239
Db 119774	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119775
OY 240	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 299
Db 119714	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119655
OY 300	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 358
Db 119654	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119655
OY 359	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 410
Db 119594	GTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 119544
RESULT 5	
AF218784	2801 bp DNA linear VRT 05-MAR-2001
LOCUS	Gallus gallus MHC Rf-Y class I alpha chain (IFV1) pseudogene,
DEFINITION	IFV1-Y-FVW7 allele, complete sequence.
ACCESSION	AF218784
VERSION	AF218784.1 GI:13195166
KEYWORDS	
SOURCE	Chickens.
ORGANISM	Gallus gallus

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS	1 (bases 1 to 134909)
TITLE	DOE Joint Genome Institute.
JOURNAL	Sequencing of Human Chromosome 5
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 134909)
TITLE	DOE Joint Genome Institute.
JOURNAL	Direct Submission
COMMENT	Submitted (07-MAR-2000) production sequencing facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. On Apr 19, 2001 this sequence version replaced gi:3256445f.


```

10970 11838: contig of 869 bp in length
11839 11839: gap of unknown length
11839 11839: contig of 628 bp in length
12547 12547: gap of unknown length
12547 12547: contig of 1422 bp in length
14049 14049: gap of unknown length
14049 14049: contig of 813 bp in length
14942 14942: gap of unknown length
14942 14942: contig of 891 bp in length
15912 15912: gap of unknown length
15912 15912: contig of 738 bp in length
16731 16731: gap of unknown length
16731 16731: contig of 1822 bp in length
18632 18632: gap of unknown length
18632 18632: contig of 1330 bp in length
20042 20042: gap of unknown length
20042 20042: contig of 1837 bp in length
21959 21959: gap of unknown length
21959 21959: contig of 743 bp in length
22040 22040: gap of unknown length
22040 22040: contig of 1087 bp in length
22863 22863: gap of unknown length
22863 22863: contig of 1353 bp in length
23950 23950: gap of unknown length
23950 23950: contig of 1765 bp in length
25462 25462: gap of unknown length
25462 25462: contig of 1397 bp in length
27227 27227: gap of unknown length
27227 27227: contig of 1946 bp in length
28704 28704: gap of unknown length
28704 28704: contig of 1338 bp in length
30730 30730: gap of unknown length
30730 30730: contig of 2149 bp in length
32148 32148: gap of unknown length
32148 32148: contig of 1962 bp in length
34457 34457: gap of unknown length
34457 34457: contig of 2458 bp in length
36419 36419: gap of unknown length
36419 36419: contig of 872 bp in length
37371 37371: gap of unknown length
37371 37371: contig of 2536 bp in length
39687 39687: gap of unknown length
39687 39687: contig of 2639 bp in length
40067 40067: gap of unknown length
40067 40067: contig of 2350 bp in length
42846 42846: gap of unknown length
42846 42846: contig of 2458 bp in length
45276 45276: gap of unknown length
45276 45276: contig of 2788 bp in length
47714 47714: gap of unknown length
47714 47714: contig of 2933 bp in length
50423 50423: gap of unknown length
50423 50423: contig of 3633 bp in length
54356 54356: gap of unknown length
54356 54356: contig of 3633 bp in length
57469 57469: gap of unknown length
57469 57469: contig of 3633 bp in length
61102 61102: gap of unknown length
61102 61102: contig of 5573 bp in length
66754 66754: gap of unknown length
66754 66754: contig of 2150 bp in length
68984 68984: gap of unknown length
68984 68984: contig of 4233 bp in length
73317 73317: gap of unknown length
73317 73317: contig of 3264 bp in length
76661 76661: gap of unknown length
76661 76661: contig of 4248 bp in length
80989 80989: gap of unknown length
80989 80989: contig of 5012 bp in length
86081 86081: gap of unknown length
86081 86081: contig of 4375 bp in length
90536 90536: gap of unknown length
90536 90536: contig of 7155 bp in length
97772 97772: gap of unknown length
97772 97772: contig of 7654 bp in length
105505 105505: gap of unknown length
105505 105505: contig of 7635 bp in length
113220 113220: gap of unknown length

```

```

113221 113300: gap of unknown length
113301 113301: gap of 9368 bp in length
122669 122748: gap of unknown length
122749 122749: contig of 20617 bp in length
143366 143445: gap of unknown length
143446 143445: contig of 606 bp in length
144051 144131: gap of unknown length
144132 144131: contig of 411 bp in length
144543 144542: gap of unknown length
144543 144542: contig of 605 bp in length
145228 145307: gap of unknown length
145308 145307: contig of 650 bp in length
145358 146037: gap of unknown length
146038 146243: contig of 206 bp in length
146244 146902: contig of 579 bp in length
146903 146982: gap of unknown length
146983 147298: contig of 316 bp in length
147299 147378: gap of unknown length
147379 148089: contig of 711 bp in length
148090 148169: gap of unknown length
148170 148859: contig of 690 bp in length
148860 148939: gap of unknown length
148940 149370: contig of 431 bp in length
149371 149450: gap of unknown length
149451 150063: contig of 613 bp in length
150064 150143: gap of unknown length
150144 150284: contig of 441 bp in length
150285 150664: gap of unknown length
150665 151269: contig of 605 bp in length
151270 151350: gap of unknown length
151350 151350: contig of 588 bp in length
151358 152017: gap of unknown length
152018 152645: contig of 628 bp in length
152646 152725: gap of unknown length
152726 153396: contig of 671 bp in length
153397 153476: gap of unknown length
153477 154126: contig of 650 bp in length
154127 154206: gap of unknown length
154207 154679: contig of 473 bp in length
154680 154759: gap of unknown length
154760 155063: contig of 304 bp in length
155064 155143: gap of unknown length
155144 155454: contig of 311 bp in length
155455 155534: gap of unknown length
155535 156214: contig of 680 bp in length
156215 156294: gap of unknown length

Query Match
Best Local Similarity 71.3%; Score 219.4; DB 2; Length 182421;
Matches 291; Conservative 0; Mismatches 116; Indels 1; Gaps 1;

OY 1 CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGGGCTGGGG 60
Db 110544 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGA 110603

OY 61 CTGGGGCTGGGGCTGGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
Db 110604 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGA 110663

OY 121 CTGGAGCTGGAGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGG 180
Db 110664 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGA 110723

OY 181 CTGAGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGG 240
Db 110724 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGA 110783

OY 241 CTNAGGCTGGAGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 300
Db 110784 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGA 110843

OY 301 CTGAGCTGGGGCTGGAGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGGG 359

```



```

repeat_region /rpt_family="(TAA)n"
complement(23395..23446)
/rpt_family="(GA)n"
repeat_region 23448..23483
/rpt_family="(CA)n"
complement(23547..23581)
/rpt_family="(TAA)n"
repeat_region complement(23582..23739)
/rpt_family="B2"
complement(23809..23949)
/rpt_family="URR1B"
repeat_region complement(23950..25874)
/rpt_family="L1_MM"
complement(25875..25927)
/rpt_family="URR1A"
repeat_region complement(26021..26112)
/rpt_family="MER45B"
repeat_region 27955..28172
/rpt_family="B4A"
repeat_region 28227..28247
/rpt_family="AT_rich"
complement(28852..28891)
/rpt_family="(CA)n"
repeat_region complement(29584..29770)
/rpt_family="B3"
complement(31305..31467)
/rpt_family="RMR15"
repeat_region complement(31573..31624)
/rpt_family="(GA)n"
repeat_region 31626..31710
/rpt_family="(CA)n"
complement(32260..32776)
/rpt_family="L1"
repeat_region 32765..33198
/rpt_family="L1_MM"
complement(33509..33739)
/rpt_family="L2"
repeat_region complement(35041..35122)
/rpt_family="B1-F"
complement(35274..35320)
/rpt_family="purine-rich"
repeat_region 35347..35424
/rpt_family="polypurine"
complement(35431..35540)
/rpt_family="RSTNE1"
repeat_region complement(35613..35797)
/rpt_family="Lx8"
complement(35863..36017)
/rpt_family="Lx7"
repeat_region complement(36150..36257)
/rpt_family="PB1D10"
complement(36634..36702)
/rpt_family="B4"
repeat_region complement(36709..36769)
/rpt_family="A-rich"
complement(36770..36921)
/rpt_family="B1_MM"
repeat_region 37088..37110
/rpt_family="(CA)n"
complement(37142..37347)
/rpt_family="B4"
repeat_region complement(37367..37470)
/rpt_family="L2"
complement(37570..37696)
/rpt_family="MLT1B"
repeat_region 37746..37836
/rpt_family="MLT1B"
complement(37865..38302)
/rpt_family="L1_MM"
repeat_region 38270..38667
/rpt_family="L1_MM"
complement(38671..38844)
/rpt_family="MLT1B"

```

```

repeat_region 39755..39896
/rpt_family="B1_MM"
repeat_region 39918..40050
/rpt_family="(GA)n"
repeat_region 40051..40181
/rpt_family="polypurine"
repeat_region 40195..40325
/rpt_family="(GA)n"
complement(40358..40593)
/rpt_family="Lx2B"
repeat_region complement(40808..40873)
/rpt_family="B4"
repeat_region complement(40875..41137)
/rpt_family="RTR1A"
repeat_region 41779..42397
/rpt_family="Lx2B"
complement(42728..43034)
/rpt_family="MYSERV"
repeat_region 43294..43443
/rpt_family="RTR1B"
complement(43449..43578)
/rpt_family="B4A"
repeat_region complement(44275..44452)
/rpt_family="B3A"
complement(44501..44695)
/rpt_family="L1MB8"
complement(45104..45256)
/rpt_family="L1_MM"
repeat_region 45752..45837
/rpt_family="L1_MM"
complement(46088..48144)
/rpt_family="L1_MM"
repeat_region 48144..49165

```

Query Match 43.0% Score 215; DB 10; Length 110312;
 Best Local Similarity 66.2% Pred. No. 8e-28;
 Matches 312; Conservative 0; Mismatches 157; Indels 2; Gaps 2;

```

QY 1 CTGAGGCTGGGGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGGGCTGGGG 60
DB 17460 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 17519
QY 61 CTGGGGCTGGGGCTGGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
DB 17520 CTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 17579
QY 121 CTGGGGCTGGAGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGGG 180
DB 17580 CTGAGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 17639
QY 181 CTGAGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 240
DB 17640 CTGAGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 17699
QY 241 CTNAGGCTGGAGCTGGGGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAG 300
DB 17700 CTGAGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 17759
QY 301 CTGAGGCTGGGGCTGGAGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAG 359
DB 17760 CTGAGGCTGGGGCTGGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGG 17819
QY 360 CTGAGGCTGGAGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 419
DB 17820 CTGAGGCTGGAGCTGGGGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 17878
QY 420 ATGAGNATGAGNCTGAGNCTGAGNCTGAGNCTGAGNCTGAGNCTGAGNCTGAG 470
DB 17879 GTCTGGGTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCT 17929

```

RESULT 14
 AC015260/c

LOCUS AC015260 7085 bp DNA linear HTG 16-NOV-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** in ordered
 pieces.
 AC015260
 VERSION AC015260.1 GI:6436075
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 7085)
 AUTHORS Adams M. and Venter, J.C.
 JOURNAL Direct Submission
 COMMENT Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 This sequence was identified as CDL10213701 by the submitter.
 For further information on this sequence e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 1. 7085
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 2102 a 1561 c 1496 g 1926 t
 ORIGIN
 Query Match 43.0%; Score 214.8; DB 2; Length 7085;
 Best Local Similarity 74.5%; Pred. No. 1e-27; Indels 0; Gaps 0;
 Matches 263; Conservative 0; Mismatches 90;
 QY 1 CTGAGCTGGGGCTGGGGCTGGGGCTGGAGCTGGAGCTGGAGCTGGGGCTGGGG 60
 DB 7077 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 7018
 QY 61 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 120
 DB 7017 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 6958
 QY 121 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 180
 DB 6957 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 6898
 QY 181 CTGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 240
 DB 6897 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 6838
 QY 241 CTNAGCTGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 300
 DB 6837 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 6778
 QY 301 CTGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 353
 DB 6777 CTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 6725

RESULT 15
 AC003574 27438 bp DNA linear INV 12-OCT-2000
 LOCUS AC003574/c
 DEFINITION Drosophila melanogaster genomic scaffold 14200001336033 section 1
 of 2 complete sequences.
 AC003574 AC003574
 ACCESSION AC003574.3 GI:10803726
 VERSION
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 27438)
 AUTHORS Adams, M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
 Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
 George, R.A., Lewis, J.E., Richards, S., Ashburner, M., Henderson, S.N.,
 Sutton, G.G., Mortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
 Brandon, R.C., Rogers, Y.H., Blazer, R.G., Chame, M., Pfeiffer, B.D.,
 Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
 Miklos, G.L., Abail, J.F., Aghayani, A., An, H.J.,
 Andrews, P., Plank, C., Baldwin, D., Bailey, R.M., Basu, A.,
 Baxendale, J., Bayraktaroglu, L., Beasly, E.M., Beeson, R.Y.,
 Benos, P.V., Berman, B.P., Bhanderi, D., Bolshakov, S., Borkov, D.,
 Botchan, M.R., Bouck, J., Brokstein, P., Brothier, P., Butts, K.C.,
 Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
 Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
 Pablo, B., Delcher, A., Deng, Z., Dey, A.D., Dew, I., Dietz, S.M.,
 Dodson, R., Dou, P., Drenth, K.J., Evans, J., Fagan, M., Garg, N.S.,
 Dunn, P., Dudin, K.J., Evangelista, C.C., Feiz, C., Fierman, S.,
 Fritschmann, W., Foster, K., Glodok, A., Gong, F., Gottlieb, J.H., Gu, Z.,
 Gelbart, W.M., Glasser, K., Glodok, A., Gong, F., Gottlieb, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
 Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wel, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
 Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C.,
 Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Paclet, J.M., Palazolo, M., Pittman, G.S., Pan, S.,
 Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kimios, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svrtk, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
 Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T.,
 Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 Zaveri, J.S., Zhang, M., Zhang, G., Zhao, J., Zhu, X., Zhu, X.H.,
 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.D., Rubin, G.M., and Venter, J.C.
 The genome sequence of Drosophila melanogaster
 Science 287 (3461), 2185-2195 (2000)
 2 (bases 1 to 27438)
 REFERENCE
 AUTHORS Adams, M.D., Celisner, S.E., Gibbs, R.A., Rubin, G.M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT On Oct 16, 2000 this sequence version replaced gi:10726352.
 FEATURES
 source
 location/Qualifiers
 1. 274388
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="X"
 join(113340..11667,12134..12342,12408..12533,12890..13377)
 /gene="CG14621"
 /product="CT34378"
 /db_xref="FlyBase:FBgn0031183"
 /db_xref="FlyBase:FBgn0031183"
 <11340..>13377
 /gene="CG14621"
 /map="19P5-19F5"
 /db_xref="FlyBase:FBgn0014621"
 /db_xref="FlyBase:FBgn0031183"
 join(113340..11667,12134..12342,12408..12533,12890..13348)
 /gene="CG14621"
 /note="CG14621 gene product"
 /codon_start=1
 /db_xref="FlyBase:FBgn0014621"
 /db_xref="FlyBase:FBgn0031183"
 /protein_id="AA050956.1"
 /db_xref="GI:7295649"
 /translation="MGKRTSGRIHAYVLLMCLFVWISSNNVIGKAVLNEFPFMT
 VTLVQCLSTILVSGPFLNMRIRKQDIPRYRYRIIVPLAIGKLAIVTSIISIMKV
 PVSVAIRKATMPLFVYIVTRFEGEKQPLVYLSLPLTGVAGTAVTETSIPMKML
 ISALISMPGSMONIFSRRKVLKNDINHLRLHLGLKLSLFTFLPLMDYMSFAVFRH

TAIKNDLVYVIALLPADGVLMNLIITAFSVLSGLVPLTAVASAKRIEYIAVSLI
 IGRNPTVWNCVGMTLAIYGVLCYRNAROLTRGROPLTPTGSPSYKYSPLLEQADPY
 YRGSVKXLSNGLHRTPGNSLLANGSGMPSGTATRLFPV"
 complement(join(14285..14495,14891..>15818))
 mRNA
 /gene="CG14615"
 /product="CT34372"
 /db_xref="FLYBASE:FBan0014615"
 /db_xref="FLYBASE:FBgn0031184"
 complement(join(14285..14495,14891..15594))
 CDS
 /gene="CG14615"
 /note="CG14615 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014615"
 /db_xref="FLYBASE:FBgn0031184"
 /protein_id="AAF50955.1"
 /db_xref="GI:7295648"
 /translation="MSDKNGDILRPLSDSEVDELIDLVKFKGIRNFHYLLYVOR
 WDRQLSEAOIPRNDLNIHSIKROFYTRRNGFRWGVYSLHDIIVOSVSPFQPDG
 AAILNCELOTOLIEWTGALITVNDLGFGRVYKELASRGVATLIDPQRCFQVSHD
 DAFCAKVPDLPSEFEIRLRADPAWHDMSGPKGSLTYLOALYRFRKSGICRSD
 TGLIIMIFQNDPSGIMLOVLPKAEERRGGLLAAMSREIARGEELITLWAVATN
 WRSEALIKRIGYKDLVNMETKLVPNSS"
 join(<16304..16313,16376..17222,17395..17607,
 17669..>17786)
 mRNA
 /gene="CG14620"
 /product="CT34377"
 /db_xref="FLYBASE:FBan0014620"
 /db_xref="FLYBASE:FBgn0031185"
 /evidence=not_experimental
 <16304..>17786
 /map="19P5-19P5"
 /db_xref="FLYBASE:FBan0014620"
 /db_xref="FLYBASE:FBgn0031185"
 /evidence=not_experimental
 join(16304..16313,16376..17222,17395..17607,17669..17786)
 CDS
 /gene="CG14620"
 /note="CG14620 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014620"
 /db_xref="FLYBASE:FBgn0031185"
 /evidence=not_experimental
 /protein_id="AAF50954.1"
 /db_xref="GI:7295647"
 /translation="AVLITEELVRKSEHNRLISTLEISLHOEDIEYIEIOWMCR
 DKILLLOSNIJARLENLHKLRLLEYLNAIINIERVENIEGESLSKDLTLNIRE
 LTVESLTCGNVRLRELVLINPCVDYRDYVATLPOLNSIDCEITPSPRLRALR
 ELSKNSIIVOVQODIENDQRIIVAKOOSALAHGAGIDEEERIKAFMOKSEH
 CPETREIRHQRHGLRERHETKSPDLKQORNLPAFCGPVNLNOKIPEPTREAD
 HYLOLEVYRHIDSLIDVQOTYTVYTKKTIPOIAYSEKVPDSEOTSCOT
 LYVNLKKLVNELLAKKSPKSPAAFPDAGKDGPEEFHGVVDISNICAPEDDP
 DLT"
 complement(join(19589..19787,19850..20031,20090..20549,
 20753..20945,21021..>21671))
 mRNA
 /gene="CG14614"
 /product="CT34371"
 /db_xref="FLYBASE:FBan0014614"
 /db_xref="FLYBASE:FBgn0031186"
 complement(join(19589..>21671))
 gene
 /gene="CG14614"
 /map="19P5-19P5"
 /db_xref="FLYBASE:FBan0014614"
 /db_xref="FLYBASE:FBgn0031186"
 complement(join(19589..19787,19850..20031,20090..20549,
 20753..20943))
 CDS
 /gene="CG14614"
 /note="CG14614 gene product"
 /codon_start=1

/db_xref="FLYBASE:FBan0014614"
 /db_xref="FLYBASE:FBgn0031186"
 /protein_id="AAF50953.2"
 /db_xref="GI:10726353"
 /translation="MSTAGKRREIYKIAWPPLYSNNMSVPRDKREFIATGPTREY
 NMVQIISIDDEPTSEFSKSPEDHPYPTKIMTPOSKCYVPLIATSGDLYRWKRG
 EPMORLECYLNKKNSNDSCAPLSPDNREVDPLNCGSSIDTPTWGLTCGRPARV
 YVAGFKTOLIAHDEYVDTAFSRAGGRMRDEASVAGDSVRFAPDLRHEHSTIYED
 PAHTALRLANKKOPNLYLVANDSCVITLIVRVCPRVATLSHRACNIGIATAP
 HSSCHITGDDPDLIMDIQMPRALEDPLATITABEGEVNDIQMGATQPMVATIC
 NKACETIRY"
 join(23250..23550,23786..23910,23973..24619,26228..26253,
 28782..29631,32524..32734,33171..33371,33378..33591,
 33655..33977,34075..34224,34281..34359,34471..34740)
 mRNA
 /gene="CG14619"
 /product="CT34376"
 /db_xref="FLYBASE:FBan0014619"
 /db_xref="FLYBASE:FBgn0031187"
 <23250..>34740
 /gene="CG14619"
 /map="19P5-19P5"
 /db_xref="FLYBASE:FBan0014619"
 /db_xref="FLYBASE:FBgn0031187"
 join(24119..24619,26228..26253,28782..29631,32524..32734,
 33171..33371,33378..33591,33655..33977,34075..34224,
 34281..34359,34471..34543)
 CDS
 /gene="CG14619"
 /note="CG14619 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014619"
 /db_xref="FLYBASE:FBgn0031187"
 /protein_id="AAF50952.1"
 /db_xref="GI:7295645"
 /translation="MRVYAPRSSSTASAGSSLYGHSAYCTSGCTSSNSGSHH
 SITGSAVDPIGGAATNGIRKISSTYDNGSLSKMPLPSSGSSNSGSGSTYITPIS
 SITSAAGSSSSNIAFRSVDPSYGTSTSTHSAGSSSSATRTSITLSGSSYDS
 PTYSRTGDLVDSMTKITTNVGLSGSKAKRDLRGGHGYVAKPPIISGVSGRKCS
 STASRDTKAVDGNKRKLSSSSSPSLARIATVSGLDIYERTSRPKNCELSRSGH
 SETDLRNGSKNETPTCTVTYLDPRWSSITRSYLRKSSDVGSSAKKILGSKYS
 ISSSSSSGVPTATFTLNSKRRNQAEVPTSYVQAGPAPAPPTPPPPPPPPPT
 NGHKKPSSNGLEAKTLNLSLSTSPKRNAIYEHNEDETDDGVSQDPAADNADVSTA

Query Match 43.0% Score 214.8; DB 3; Length 274388;
 Best local similarity 75.2% Pred. No. 8.2e-28;
 Matches 261; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1 CTGAGGCTGGGGCTGGGGCTGAGCTGGAAGCTGAGGCTGAGGCTGGGGCTGGGG 60
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 178082 CTGGAGCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGG 178023
 QY 61 CTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 178022 CTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGG 177963
 QY 121 CTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 177962 CTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGG 177903
 QY 181 CTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 177902 CTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGG 177843
 QY 241 CTMAGGCTGAGGCTTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 177842 CTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGG 177783
 QY 301 CTGAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 347
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 177782 CTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGG 177736

Search completed: October 10, 2002, 19:01:38
 Job time : 1301.87 secs

Fri Oct 11 09:30:50 2002

us-09-489-101a-17_copy_1_500.rge

Page 17

[illegible]

[illegible]

```

RESULT 2
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURES:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

```

[illegible]

Db 3937747 ATGCGCGCGGCGCGCTTGCCCGCGGTGCGCTTGCGCGCAGC 3937703

```

1      RESULT 3
2      US-09-197-649-7/c
3      : Sequence 7, Application US/09197649
4      : Patent No. 6194550
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Gold, Larry
9      : APPLICANT: Therk, Craig
10     : APPLICANT: Pritnow, David
11     : APPLICANT: Smith, Jonathan D.
12     :
13     : TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
14     :
15     : FILE REFERENCE: NEX02/C1-CON
16     :
17     : CURRENT APPLICATION NUMBER: US/09/197,649
18     :
19     : EARLIER FILING DATE: 1998-11-23
20     : EARLIER APPLICATION NUMBER: 07/829,461
21     :
22     : EARLIER FILING DATE: 1992-01-31
23     : EARLIER APPLICATION NUMBER: 07/739,055
24     :
25     : EARLIER FILING DATE: 1991-08-01
26     : EARLIER APPLICATION NUMBER: 07/561,968
27     :
28     : EARLIER FILING DATE: 1990-08-02
29     : NUMBER OF SEQ ID NOS: 26
30     :
31     : SOFTWARE: PatentIn Ver. 2.0
32     :
33     : SEQ ID NO 7
34     :
35     : LENGTH: 390
36     :
37     : TYPE: DNA
38     :
39     : ORGANISM: Artificial Sequence
40     :
41     : FEATURE:
42     :
43     : OTHER INFORMATION: Description of Artificial Sequence: Sequence
44     :
45     : OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
46     :
47     : OTHER INFORMATION: fragments having NcoI restriction sites.
48     :
49     : US-09-197-649-7

```

[illegible]

Fri Oct 11 09:30:47 2002

us-09-489-101a-16_copy_1_500.rni

Page 3

```

1  APPLICANT: Sutcliffe, Thomas D.
2  APPLICANT: Rodriguez, Raymond L.
3  TITLE OF INVENTION: Production of Mature Proteins
4  TITLE OF INVENTION: In Plants
5  NUMBER OF SEQUENCES: 23
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: Dehlinger & Associates
8  STREET: 350 Cambridge Ave., Suite 250
9  CITY: Palo Alto
10 STATE: CA
11 COUNTRY: USA
12 ZIP: 94306
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Diskette
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: DOS
17 SOFTWARE: FastSeq for Windows Version 2.0
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/023.173
20 FILING DATE: 13-FEB-1998
21 CLASSIFICATION: 435
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 60/038.168
24 FILING DATE: 13-FEB-1997
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Pelithory, Joanne R
27 REGISTRATION NUMBER: P4295
28 REFERENCE/DOCKET NUMBER: 0665-0007.30
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 650-324-0880
31 TELEFAX: 650-324-0960
32 INFORMATION FOR SEQ. ID NO.: 4:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 1140 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 IMMEDIATE SOURCE:
39 CLONE: Codon-optimized Rany3D signal-prosubtilisin BPN'
40 US-09-023-173-4

```

[illegible]

Oy	425	CGCACCCAGGGAATGGCCAGAGTGTGCACAGGGCATCTGCCAACCAAGGACGCCGCCA	484
Db	690	CCCGAGCGGCGAGCGCGCCCTCAAGGCGCGCCTTGACAAAGGCGCTTCGCCAGCGGGTCT	749
Oy	485	GGGCATTCGCC	494
Db	750	CGTCGTCCGC	759

RESULT 5
 US-09-056-556-176
 ; Sequence 176, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasar A.W.
 APPLICANT: Dillon, Davin C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 NUMBER OF SEQUENCES: 241
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-Apr-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.457
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 176:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 494 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

	Query Match Best Local Similarity	15.4% 49.6%	Score 77; Pred. No. 3.8e-06;	DB 4; Length 494;	Mismatches 197; Conservative 0;	Indels 200;	Gaps 0;
QY	38	CGGTATGAGAGCGCCGACAGGCGATCGCTAATGAGAGACACACCCAGTCATGCCCA	97				
DB	16	CGCGAAGGGCGCGGACGCTGTGGCAGCGCTGAGAGCGCTGGCGSACAAAGTTCGTGAC	75				
QY	98	CGAGGAAGCGCCCGACGGCGATCGCCGAGAGAGCCATTCAGGCGCATTCGCAAGAGAGGT	157				
DB	76	CGAAGCGCGACGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	133				
QY	158	TGCCCAAGGCGATCTCCCAATGTGGTGTGGCGCACAGGCGATCTGCATATGAGAGGCCACCA	211				
DB	136	CGCGAAGGGCGCGCTTCGCGCGCGACGCGCGGTACAGGGCGGCCCAACGCGCGCGCTACGT	193				
QY	218	GGGCATTCGCGAATCTGAGACGCGCTTCATGGCTTTCGCCAACGGGAGCGCGCTCTCAAGCTT	277				
DB	196	CGCGACACCGCGCGCTGTGTGGCGCGCGCGCACAGCGGTGTGTGGCGCGCGCGCGCGCGCGCT	255				
QY	278	CGCGAAGCGGGAGCGCGCGCCAGGCGATCGCCACAGGGAGCGCGACCAAGGCGATGSGCGCA	331				
DB	256	CTTTTGGCGGTGCT	315				

Db 1417 A 1417

RESULT 8
US-09-105-537-7
: Sequence 7, Application US/09105537A
: Patent No. 6265202
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600,438051
: CURRENT APPLICATION NUMBER: US/09/105,537A
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 7
: LENGTH: 1248
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-105-537-7

Query Match 14.9% Score 74.4; DB 4; Length 1248;
Best Local Similarity 49.5%; Pred. No. 1,le-05;
Matches 220; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 59 GGGGATCGCTAATGAGACACACCCAGTGCATCGCCAAACGAGAGCCGCCAGGGCAT 118
Db 444 GGGGATCGCTAATGAGACACACCCAGTGCATCGCCAAACGAGAGCCGCCAGGGCAT 118
QY 119 CGCCGAGAGCGCCATCCAGGACATCGCCAAACGAGAGGTTGCCAGGGCATCGCCAAATGG 178
Db 504 GATCTTCGAGCGCCGAGCGCCCTCGGCTGCGGCGGTCGAGCGCGCCGCCGCGGACCGCT 563
QY 179 GGTGCGCCAGACGAGGATCGCCAAATGAGAGGAGGAGGAGGATCGCCAAATGGAGAGC 238
Db 564 CGCGGAGCGCGGAGGTCTTCACTTCACGACGACCAAGGCGGCTTCGAGAGCGCGG 623
QY 239 CGTTCACGCGCTTCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 295
Db 624 CGCGGAGCGCGGAGGTCTTCACTTCACGACGACCAAGGCGGCTTCGAGAGCGCGG 623
QY 296 CGAGGAGATCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355
Db 684 CTTCGAGCGCTTCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 743
QY 356 CATCGCTAACGAGAGCGCGCTTCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 415
Db 744 CGCGGAGCGCGGAGGTCTTCACTTCACGACGACCAAGGCGGCTTCGAGAGCGCGG 803
QY 416 CAACGAGAGCG 475
Db 804 CAACGAGAGCG 863
QY 476 CG 499
Db 864 CGACCG 887

RESULT 9
US-09-056-556-167/C
: Sequence 167, Application US/09056556
: Patent No. 6350456
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Dillon, David C.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
: NUMBER OF SEQUENCES: 241
: CORRESPONDENCE ADDRESS:

TREATY

ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 167:
SEQUENCE CHARACTERISTICS:
LENGTH: 1439 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-167

Query Match 14.9% Score 74.4; DB 4; Length 1439;
Best Local Similarity 47.6%; Pred. No. 1,le-05;
Matches 219; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 8 CGGATTCCTGAGAGCGCTTCGCCAGGCGATCCCTAATGAGAGCGCCGACCGGCGATCCG 67
Db 1413 CGGATTCCTGAGAGCGCTTCGCCAGGCGATCCCTAATGAGAGCGCCGACCGGCGG 1354
QY 68 TAATGAGAGCAGCAGCGGATGATCGCCAAAGAGAGAGCGCGCGCGCGCGCGCGCGCG 127
Db 1353 TGTGCGCTTCG 1294
QY 128 CGGATTCAGGAGCGATTCGCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 187
Db 1293 TGTGCGCTTCG 1234
QY 188 ACAGGAGATTCGCAATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 247
Db 1233 TGTGCGCTTCG 1174
QY 248 CTTCGCGCAAGGAGAGCG 307
Db 1173 TGTGCG 1114
QY 308 CAACGAGAGCG 367
Db 1113 CGCGGAGCG 1054
QY 368 GAGCGCGCTTCAGAGGAGATTCGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 427
Db 1053 CGCGGAGCG 994
QY 428 CGCGGAGGAGATTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 467
Db 993 TGTGCGCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 954

RESULT 10
US-09-105-537-3/C
: Sequence 3, Application US/09105537A
: Patent No. 6265202
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhuo, L.
TITLE OF INVENTION: DNA encoding methymycin and plikromycin
FILE REFERENCE: 600 438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 3
LENGTH: 13613
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-3

Query Match 14.9%; Score 74.4; DB 4; Length 13613;
Best Local Similarity 49.5%; Pred. No. 1.1e-05;
Matches 220; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 59 GGGCATGCTATAGAGACACCAAGTGCATGCCAAGAGAGAAAGCCGAGGCGCAT 118
DB 12440 GGGCCGCTCTGCGCCGCGGACAGCTGGAGAGTGGCCGAGAGAGCGGCTGCGCT 12381
QY 119 CGCCGAGAGAGCCATCCAGGCGATGCCAAGAGAGAGTGGCCGAGGCGCATGCGCATG 178
DB 12380 GTACTTGCAGCGCCGCGACGCTCGGCTGCGCGGTGAGAGCGCGCGCGCGAGGCT 12321
QY 179 GGTGCGCGACAGGAGGATGCCAATGAGAGAGAGCCAGGCGATGCCAATGAGAGAG 238
DB 12320 CGGCGAGCGGAGGCTTCAAGCTTCCAGCCAGGAGCGCTTCAAGCGCTTCAAGCGCG 12261
QY 239 CGTCCAGCGGTGTGGCCAGAGGGA---GCGCGTCTTCAAGCTTCCAGGAGAGCGCGC 295
DB 12260 CGCGGTGTACCGAGAGAGAGCGAGCTCGCGCGCGAGATCGCGCGCTTCAAGCTTGG 12201
QY 236 CCAGGCGATGCCAAGAGGAGCGCCAGGAGGATGAGGAGGATGCCAATGAGTCCAGG 355
DB 12200 CTTCGACTCTCGCGGCGAGCGCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 12141
QY 356 CATGCTTACAGAGAGCGCGCTCCAGGAGATGCTAAGAGAGTGGCGCGCGAGGCGATCGC 415
DB 12140 CGCGCGATGAGGCTTCACTCTCCAGCGCTTCCGAGAGTCAAGAGGAGGAGGAGGAGG 12081
QY 416 CAACGAGAGCGCGCGCCAGGAGGATGCCAGAGATGCGCAGAGGAGTCCAGGAGAGAG 475
DB 12080 CAACGAGCGCGCTTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 12021
QY 476 CGCGCGCGAGGCGATGCCAAGAA 499
DB 12020 CGACCGCGCAGCGGCTCAACAGCA 11997

RESULT 11
US-09-177-349-2/c
Sequence 2, Application US/09177349
Patent No. 6268201
GENERAL INFORMATION:
APPLICANT: Alland, David
APPLICANT: Bloom, Barry R.
APPLICANT: Jacobs, Jr., William R.
TITLE OF INVENTION: 11B INHA AND INIC GENES OF MYCOBACTERIA AND METHODS
FILE REFERENCE: 96700/491
CURRENT APPLICATION NUMBER: US/09/177,349
CURRENT FILING DATE: 1998-10-23
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 5036
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-177-349-2

Query Match 14.8%; Score 74; DB 4; Length 5036;
Best Local Similarity 48.2%; Pred. No. 1.3e-05;
Matches 209; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 7 ACGCATTTCTGAGAGACCTGCCCCAGGCGCATGCTTAATGAGAGAGCCGAGGCGCATG 66
DB 935 ACTCGTGTGCTGAGACCCCGGCTGAGCGCGCTGACATTTGCGAGCGCGCGAGCCGACCA 876
QY 67 CTAAATGAGACACCCAGTGCATGCGCAACAGAGAGAGCCGCGCCAGGCGATGCGCGAG 126
DB 875 CCGGCTGAGACACCGATGCTGATGCTGATGCTGCGCGCTGCGCGCGCGCGCGCGCG 816
QY 127 ACGCATTCAGGCGCATGCGCAACAGAGAGGTTGCGCCAGGCGATGCGCAATGAGGCGCG 186
DB 815 CCAGCGCGCAAGGCG 756
QY 187 CACAGGCGATGCGCATGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
DB 755 CCGAGCG 696
QY 247 GCTTCGCCAAGGAGAGCG 306
DB 695 CCAGAGCG 636
QY 307 CCAGCGCGAGCG 366
DB 635 CCATTAACCTTAACCGAGCG 576
QY 367 AGAGAGCG 426
DB 575 CCAGAGCG 516
QY 427 CCGCGCGAGGAGATC 440
DB 515 AGAGCGAGCG 502

RESULT 12
US-09-320-878-21/c
Sequence 21, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 21
LENGTH: 5970
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-21

Query Match 14.8%; Score 74; DB 3; Length 5970;
Best Local Similarity 49.3%; Pred. No. 1.3e-05;

Matches 219; Conservative 1; Mismatches 221; Indels 3; Gaps 1;

```

OY 59 GGGCATTGTAATGAGACACACCACCACTGATGCGACAGAGAGAGCCGCCAGGCAAT 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 GGGCGGCGCCCTGGCGCGGACACCTCTGGAGAGTGGCCGAGACGACGACCTTCGCT 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 119 CGCGAGAGAGCCATCAGAGGCGATCGCCACAGAGAGTGTCCAGGCGATCGCCAGTGG 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 GTACTTGGAGCGCGCGACAGCCCTCGGCTGCGGTGAGAGCGCGCGCGCGAGGCT 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 179 GGTGGCGGACAGAGGCGATCGCCATGAGAGAGCGACAGGAGGATCGCCACATGAGAGC 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 CGGAGAGCGCGAGGCTTTCAGTTCCAGCCACCAAGCGCGCTTCAAGCCTTGAAGGGGG 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 239 GCTCCAGCGCTTGGCCAGAGGGA---CGCGTCTAGTGTGCGACAGGAGGAGCGCGC 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 CGCGGTGTACGAGAGCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 296 CGAGGCGATCGCCACAGGCGGAGCGCCACAGAGGATGGGACAGAGTCACTCCACGCG 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 CTTCGACCTGCGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 356 CATCGCTAAGAGAGAGCGCGCTCAGAGGATGCTAAGAGAGTGGCGCGCGCGCGATCGC 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 CGCGCGCATGGGCTCCTCCTCCGACGCGCTTCCGAGATGATGACCGGAGCGCGCG 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 416 CAACGAGAGCGCGCGCGAGAGATGCGCGAGATGTCGACAGAGGATCGCCACAGAGGA 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CAACGAGCGCGCTACCGGAGACCTCGCGGACCTCCCGCGCGCTCGTGTGCGCGACCA 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 476 CGCGCGCGAGGCGCGCGCAACA 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 CGACCGCGACGCGCTCAACACCA 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13
US-08-387-942C-1
; Sequence 1, Application US/08387942C
; Patent No. 5939289

```

; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELEGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12588 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Azotobacter vinelandii
; STRAIN: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 290..1951
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2227..6438
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6702..9695
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9973..12588
; US-08-387-942C-1

```

Query Match 14.78; Score 73.4; DB 2; Length 12588;
Best Local Similarity 46.8%; Pred. No. 1.7e-05;
Matches 230; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

```

OY 4 TCTAGCGCATTTGCTGAGAGCGCTGCGCCAGGCGATCGCTAATGAGAGCGCGCACAGGGA 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11237 TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 64 TCGCTAATGAGAGACACACCGCATGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCG 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11297 TCTCTGCGCGCGGAGAGACACCGCATGCGTACCGCGCGCGCGCGCGCGCGCGCGCG 11356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 124 AGAGCGCATTCAGAGGCGATCGCCACAGAGAGGTTGCCAGAGCGCGCGCGCGCGCGCG 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11357 ACTTGAGCGCAGCGAGACCGCATGCGTCTCGCGCTGCGCTTACCGCGCTGCGCGA 11416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 184 CGCGACAGGCGATCGCATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11417 ACGGTTTGAAGCGCACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 244 ACGGTTTGCACACAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11477 AGACCTTACAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 304 TCGCCACAGGCGGAGCGCGACAGAGGCGATGGGCAAGAGGTACCATCATCGGATCGCTA 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11537 TCGGCGAGTTCAACAGATGCGCAACCTGTTGTGAGCGCGCGCGCGCGCGCGCGCGCG 11596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 364 ACGAGAGCGCGCGTCCAGGCGATGCGCTAACAGAGGTGCGCGCGCGCGCGCGCGCGCGCG 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11597 GCACGACACACTGTCCGCGACGATGCGCGGAGAACCTCTCTGCGGTACGCGGCGAG 11656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 424 ACGCGCGCGCGAGATTCGCGAGAGTGTGCGACAGGCGATGCGCCACAGAGAGCGCGCG 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11657 ACACCTTACAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 484 AGGCGATCGCC 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11717 TGACCGCGCGC 11727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pihromycin
; FILE REFERENCE: 600.438051

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 seconds
(without alignments)
9699,805 Million cell updates/sec

Title: US-09-489-101a-16_copy_1_500

Sequence: 1 gagctacgcatgctgtag.....cccaaggcatgcgaacaag 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estda:*
2: em_esthm:*
3: em_estlin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	153.4	30.7	697	9	AL110383 DKF2P434E
2	124.6	24.9	179	9	AL041090
3	118.6	23.7	758	10	B1829297 603079427
4	113	22.6	673	10	B1561135 603253662
5	98.6	19.7	582	10	BG724320 602697968
6	94.2	18.8	241	9	A1904151 CM-BT043-
7	85.4	17.1	780	10	BG847089 102401580
8	80.8	16.2	822	12	CNS04SGK
9	79.6	15.9	640	10	BE427000 PSR6561 I
10	79.2	15.8	884	10	BE417341 HVSMEK001
11	76.4	15.7	920	9	BE216934 EST0477 T
12	76.6	15.3	602	9	AU176645
13	74.8	15.0	681	12	CNS02EOD
14	73.8	14.8	915	12	CNS03P82
15	73.4	14.7	401	9	AV438522
16	73.4	14.7	511	10	BG739444 EMI_81_A0
17	70.2	14.0	511	10	BG773077 602721380

18	69.8	14.0	454	9	BE125156
19	69.6	13.9	673	10	BG907710
20	69.2	13.8	538	10	BM178969
21	69.2	13.8	679	9	AL506812
22	69.2	13.8	814	10	BE260740
23	68.6	13.7	554	10	BG367845
24	68	13.6	530	10	BE483260
25	68	13.6	689	9	AU076299
26	67.8	13.6	619	10	BG300721
27	67.6	13.5	564	10	BE619259
28	67.6	13.5	652	10	BG314481
29	67.4	13.5	938	9	BG874723
30	67.4	13.5	938	9	BE213813
31	67.2	13.4	625	10	BG906607
32	67.2	13.4	925	12	CNS0092P
33	67.2	13.4	704	9	AU088578
34	66.8	13.4	461	10	BE317259
35	66.8	13.4	658	9	AU088571
36	66.6	13.3	456	9	AU022296
37	66.6	13.3	529	10	BE474993
38	66.6	13.3	585	10	BE470936
39	66.6	13.3	676	9	AV916390
40	66.4	13.3	633	9	AV941861
41	66.2	13.2	596	10	BE360540
42	66.2	13.2	770	10	BE265823
43	66	13.2	459	10	BE605180
44	65.6	13.1	401	9	AV438522
45	65.6	13.1	537	10	BG904964

ALIGNMENTS

RESULT 1
LOCUS AL110383
DEFINITION DKF2P434E0731.t1 434 (synonym: htes3) Homo sapiens cDNA clone
VERSION AL110383
KEYWORDS DKF2P434E0731 5', mRNA sequence.
SOURCE EST.
ORGANISM human.
REFERENCE Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE EST (Blum, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Blum H

MPSS
Am Kiofepfepitz 18a D-82153 Martinsried, Germany
This is the S. sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No si sequence available.
This clone (DKF2P434E0731) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2P434E0731"
/clone.lib="434 (synonym: htes3)"
/tissue.type="testis"
/dev.stage="adult"
/lab.host="DH10B"
/note="vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 217 a 181 c 180 g 119 t

Db 220 CTGACGACTCTGACGACGCTCCGGCTGTGAGAAAGGCGGCCCTCGCAAGGCGCTCGCA 279

Qy 76 AC 135

Db 280 AGAAG 339

Qy 136 AGGAGCATCCGCAACAG 195

Db 340 ACAGAGCCCCGAAAG 399

Qy 196 TCCGCAATGAG 255

Db 400 GCGACACACATGACGACGACGATGACGACGACGATGACGACGACGATGACGACGACGAT 459

Qy 256 ACAGGAG 315

Db 460 AGGAG 519

Qy 316 ACAGGAG 375

Db 520 AGT 576

Qy 376 TCCAGGAG 435

Db 577 ---AGGAG 633

Qy 436 GAATGCGGAG 492

Db 634 ACAGACGAGT 690

RESULT 8
CNS04SGC/c
LOCUS
DEFINITION
Tetradodon nigroviridis genome survey sequence PUC-ori end of clone 007M14 of library H from Tetradodon nigroviridis, genomic survey sequence.

ACCESSION
AL305165.1
GI:8196096
KEYWORDS
GSS: genome survey sequence.
SOURCE
Tetradodon nigroviridis.
ORGANISM
Tetradodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetradodon.

REFERENCE
1 (bases 1 to 822)
Roest-Crollius H., Jallion O., Dasilva C., Fizames C., Fisher C., Bouneau L., Billault A., Quetier F., Sautin W., Bernot A. and Weissenbach J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis

TITLE
Unpublished
2 (bases 1 to 822)
Roest-Crollius H., Jallion O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brothier P., Quetier F., Sautin W. and Weissenbach J.
Human gene number estimate provided by genome wide analysis using Tetradodon nigroviridis DNA sequence

JOURNAL
Unpublished
3 (bases 1 to 822)
Genoscope.

REFERENCE
Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetradodon>.

FEATURES
Source
1..822
/organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone="007M14"

/clone_11b="H"
/note="Genoscope sequence ID : COBH007BG07XEL-end : PUC-ori"

BASE COUNT 147 a 262 c 203 g 209 t 1 others

ORIGIN

Query Match 16.28; Score 80.8; DB 12; Length 822;
Best Similarity 50.98; Pred. No. 0.021; Indels 3; Gaps 1;
Matches 217; Conservative 0; Mismatches 206;

Qy 20 GAGAGT 79

Db 556 GAGT 497

Qy 80 CACAGT 136

Db 496 AGGATGT 437

Qy 137 GAGATGT 196

Db 436 AGGT 377

Qy 197 GAGATGT 256

Db 376 GGT 317

Qy 257 GAGATGT 316

Db 316 GGT 257

Qy 317 GGT 376

Db 256 GGT 197

Qy 377 CAGAGT 436

Db 196 CTAAGT 137

Qy 437 AATGCG 442

Db 136 TGACGC 131

RESULT 9
LOCUS
DEFINITION
PSR6561 ITPC PSR wheat Endospem library Triticum aestivum cDNA

ACCESSION
BE427000
BE427000
KEYWORDS
EST
SOURCE
EST
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.

REFERENCE
1 (bases 1 to 640)
Anderson O.A., Appels R., Bailey P., Blake T., Close T., Cloutier S., Dubcovsky J., Feuillet C., Gale M., Graner A., Gustafson P., Hermann R.G., Holton T., Jacquemelin J.M., Jia J., Joudrier P., Langridge P., Lazo G.R., Lin J.J., McGuire P., Ogihara Y., Pecchioni N., Qualset C., Schuch W., Selvaraj G., Shariflou M., Sorrells M., Warburton M. and Wenzel G.
International Triticaceae EST Cooperative (ITPC): Production of Expressed Sequence Tags for Species of the Triticeae

TITLE
Unpublished (2000)
Contact: Bailey P
Cereals Group, John Innes Centre
Norfolk, Norwich NR4 7UH UNITED KINGDOM
Tel: 44 1603 452571 ext. 2587
Fax: 44 1603 502241
Email: paul.bailey@jic.ac.uk
International Triticaceae EST Cooperative (ITPC)

JOURNAL
COMMENT

255 AACGGAGAGCGCTCTCTGAGCTTGGCCAGAGGAGCCGCCAGGAGCATGCCCAACGG 314
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 348 AACCTTGGGAGCGCCCTCGCCGATCCGACAGGAGCGCCCATGATCCGACACAGGGG 407
 315 GAGCCGACCAAGGAGCATGGGCAACAGAGGTCACTCCAGGATCCGTAACGAGAGCGCC 374
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 408 GAGCCGAGCATCGCTCAATGCTGCGAGGCGCTCCGACAGCTCGCTCATGGGCGAC 467
 375 GTCCAGGAGCATGCTCAACAGAGGTGGCCGCGGAGCATGCGCAACAGAGAGCGCCGAG 434
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 468 GTCCGCTGCTCGCGAGCATGAGAGCAGAGGATTTCTCATGCAACAGCATGCGC 527
 435 GGATCGCCGAGGATGCTCCGACAGCA 462
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 528 GACCTTACGACCTGCTCATGACAGCA 555

RESULT 11
 BE216934 990 bp mRNA linear EST 03-JUL-2000
 LOCUS BE216934
 DEFINITION EST0477 Triticum aestivum Lambda Zap Triticum aestivum cDNA clone
 JAL1_5B_H05_T3 5', mRNA sequence.
 ACCESSION BE216934
 VERSION BE216934.1 GI:8904620
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Triticum.
 1 (bases 1 to 990)
 REFERENCE Anderson, J.M., Williams, C.E. and Goodwin, S.B.
 TITLE Analysis of an EST database reveals a probable Cf2 resistance gene
 homolog in wheat
 JOURNAL Unpublished (2000)
 COMMENT Contact: Anderson, J.M.
 Crop Production & Pest Control Research Unit
 USDA-ARS
 1150 Lilly Hall, West Lafayette, IN 47907, USA
 Tel: 765-494-5565
 Fax: 765-496-2926
 Email: janderson@purdue.edu
 Seq primer: T3
 High quality sequence stop: 990.
 Location/Qualifiers
 1..990
 /organism="Triticum aestivum"
 /strain="P29"
 /db_xref="taxon:4565"
 /clone="JAL_5B_H05_T3"
 /tissue_type="leaf"
 /dev_stage="9 day old seedlings"
 BASE COUNT 182 a 360 c 290 g 158 t
 ORIGIN

Query Match 15.7%: Score 78.4; DB 9; Length 990;
 Best Local Similarity 50.0%; Pred. No. 0.043;
 Matches 230; Conservative 0; Mismatches 221; Indels 9; Gaps 1;

3 GTCTACGGCATTTGCTGAGAGCGTCCGAGGAGCATGCTATATGAGAGCGGACGAGGC 62
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 141 GCTATATTCGGCGCCAGGCGCGCGCGCGCGCATGTCGACCGGCGCTCATCGCGAA 200
 63 ATCGCTATATGAGAGCAACCCAGTGCATCCGCAAGAGAGCGGCGGAGCATCGGC 122
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 201 ATCAAGCGCGCGCTGACATCCAGTCATGAGGCGGCGGAGCATCGGCGCTGTCAG 260
 123 GAGAGCGCATTCAGGAGCATTCGCAAGAGAGGTTGCCCGAGGAGCATTCGCAATGGG 182
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 261 GCCCGATTCCTGAGTCATCGGCGTGTGACTGAGAGAGC-----GAGGTC 311
 183 GCCGCAAGGAGCATGCCAATGAGAGCGCAAGGAGGAGCATGCCAATGGAGCGCGTC 242

312 CTCAGCTCCGAGAGAGCGCCGACCATCATCAACAGACACTTCCGCTCCCTTCCGTC 371
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 243 CAGGCTTTCGCAAGGAGGAGCGCTGCTGAGTTCGTCGCAAGGAGAGCGCCGAGGCG 302
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 372 TCGCGTCCGCAACCTGGGCGAGGCGCTCGCGCATTCGCGAGGAGCGCGCATGTGTC 431
 303 ATCCCAACGAGGAGCGCACAGAGGAGCATGGGCAACAGAGGTCACTGACGAGCATGCT 362
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 432 CGCACAAAGGAGCGAGCGCGGACCGCATGTATCATGAGCGCTCCGACGCTCCGCTCC 491
 363 AACGAGAGCGCGTCCAGGAGCATGCTAAGCAGAGTGGCCGCCAGGAGCATTCGCAAGAG 422
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 492 GTCATGGGCGAGCTGGGCTGCTCCGCGAGCATGAGACGACGAGGAGTATCATGCTGCC 551
 423 GAGCGCGCCGAGGAGATGCGCGAGGATGTCGACAGGAGCA 462
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 552 AAGAGCATCGCGCGACCATGACGACCTGCTCATGAGACCA 591

RESULT 12
 AUI76645 602 bp mRNA linear EST 21-MAR-2001
 LOCUS AUI76645
 DEFINITION AUI76645 Sugano-Kawakami 5' end enriched cDNA library (Ola) from
 HNI Oryzias latipes cDNA clone Ola14.06c similar to pIRIS2641
 hydroxyproline-rich glycoprotein - maize, mRNA sequence.
 ACCESSION AUI76645
 VERSION AUI76645.1 GI:13424771
 KEYWORDS EST.
 SOURCE Japanese medaka.
 ORGANISM Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Altheriomorpha;
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 602)
 REFERENCE Naruse, K., Mitani, H. and Tanaka, M.
 TITLE Medaka EST Project in University of Tokyo (2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Kiyoshi Naruse
 Department of Biological Sciences
 Graduate School of Science, University of Tokyo
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
 Tel: 81-3-5841-4443
 Fax: 81-3-5841-4410
 Email: naruse@biol.s.u-tokyo.ac.jp
 This clone was isolated from Sugano-Kawakami 5' end enriched cDNA
 library (Ola).
 Location/Qualifiers
 1..602
 /organism="Oryzias latipes"
 /strain="HNI"
 /db_xref="taxon:8090"
 /clone="Ola14.06c"
 /tissue_type="Sugano-Kawakami 5' end enriched cDNA library
 (Ola) from HNI"
 /dev_stage="adult"
 BASE COUNT 132 a 246 c 175 g 49 t
 ORIGIN

Query Match 15.3%: Score 76.6; DB 9; Length 602;
 Best Local Similarity 49.5%; Pred. No. 0.078;
 Matches 227; Conservative 0; Mismatches 229; Indels 3; Gaps 1;

29 CCAGGAGCATGCTATATGAGAGCGCGAGGAGCATGCTATATGAGAGCAACGACGAGTG 88
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 119 CCAAGCTGTGGCGAGCGCGCGCAAGCAAGCATGTGGCGAGCGCGCGCAAGCAAGAGC 178
 89 CATGCCCAAGAGAGAGCGCGCGAGGAGCATGCGCGAGAGCGC---CATTCAGGAGCATGCGC 145
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 179 TGTGGCGAGCGCGCGCGAGCAACGAGCTGTGGCGAGCGCGCGAGCAACGAGCTGTGGC 238

QY	146	CAACGAGAGAGTGGCCAGGAGCATGCGCCAACTGGGGTGCGCCACAGGGCACTTCGCACATATA	205
Db	239	CGAGCGCCGCGCAACCAAGACTGTGGCGAGCGCCGACCAACCAAGACTGTGGCCGACGC	298
QY	206	GGAGCGCCACCAAGGAGCATGCGCAACTGGGGAGCGCGCTGTCCACGGCTTCGCCAAAGGGGACGC	265
Db	299	CGCCAGCAACCAAGACTGTGGCGAGCGCGCCGACCAACCAAGACTGTGGCCGACGCGCCGACG	358
QY	266	CGTCTTCAAGCTTGGCCAAAGGGAGCGCGCCGACGAGGCACTCCCAACGGGGAGCCCAACAA	325
Db	359	CAACCAAGACTGTGGCGAGCGCGCCGACCAACCAAGACTGTGGCCGACGCGCCGACCAACCA	418
QY	326	GGGCACTGGGCAAGSAGSTCAACCTCCAGCGCATGCTCAAGAGGAGGCGCTGTCCAGGGGCAAT	385
Db	419	GACTGTGGCGAGCGCGCCAGCAACCAAGACTGTGGCGAGCGCGCCGACCAACCAAGACTGT	478
QY	386	CGCTTCAAGSAGSTGGCGCGCCAGGCGCATGCGCCAAAGSAGSAGCGCGCTCCAGGGAAATGCCCCA	445
Db	479	GGCGCAAGCGCGCCAGCAACCAAGACTGTGGCGAGCGCGCCGACCAACCAAGACTGTGGCGCA	538
QY	446	GGATGTGGCAAGGGGATGCCCAAGSAGSAGCGGCCCA	484
Db	539	CGCGCGCAGCAACCAAGCTGTGGCGAGCGCGCCGACCAAGCA	577

RESULT 13	
CNS02EOD/c	
LOCUS	681 bp DNA linear GSS 13-MAY-2000
DEFINITION	Tetradon nigroviridis genome survey sequence r7 end of clone 262H14 of library G from Tetradon nigroviridis, genomic survey sequence.

ACCESSION	AL193990
VERSION	AL193990.1
KEYWORDS	GSS: genome survey sequence
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis

REFERENCE	AUTHORS
1 (bases 1 to 681)	
Roest-Crollius, H.,	Jallou, O.,
Dasilva, C.,	Fizames, C.,
Fisher, C.,	

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 681)
AUTHORS	Roest-Collins, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,

TITLE	Human gene number estimate provided by genome wide analysis using
JOURNAL	Tetradon nigroviridis DNA sequence
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 681)
GENEBANK	Genbank

REMARKS
JOURNAL
COMMENT
Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigriviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES	location/Qualifiers
source	1..681
	/organism="Tetradodon nigroviridis"
	/db_xref="taxon:99883"
	/clone="262H14"
	/clone_1bp="G"
BASE COUNT	/note="Genoscope sequence ID : C0AG262D07LPI-end : T7"
ORIGIN	74 a 105 c 170 g 282 t 50 others

Query Match 15.0%; Score 74.8; DB 12; Length 681;

Best Local Similarity 45.4%; Pred. No. 0.13;
Matches 173; Conservative 21; Mismatches 187; Indels 0; Gaps 0;

[illegible]

RESULT 14		
CNS03P82	915 bp	DNA linear
LOCUS		
DEFINITION		
CNS03P82	915 bp	DNA linear
		GSS 17-MAY-2000
		Tetradon nigroviridis genome survey sequence PUC-ori end of clone
		04AF19 of library G from Tetradon nigroviridis, genomic survey

ACCESSION	AL254315
VERSION	AL254315.1
KEYWORDS	GI:7975327
SOURCE	GSS: genome survey sequence.
ORGANISM	Tetradodon nigroviridis.
	Tetradodon nigroviridis.

REFERENCE
AUTHORS
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
1 (bases 1 to 915)
Tetraodontidae; Tetraodon.
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

TITLE	JOURNAL	REFERENCE
Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	unpublished	2 (bases 1 to 915)

TITLE	JOURNAL	REFERENCE
Saurin, W. and Weissenbach, J. Human gene number estimate provided by genome wide analysis using Tetraodon nigriviridis DNA sequence	Unpublished	3 (bases 1 to 915)

AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.crs.fr/tetraodon .
FEATURES	Location/Qualifiers
source	1..915

```
/organism="Tetraodon nigroviridis"  
/db_xref="taxon:99883"
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 84.9333 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-16_COPY_1_500
Perfect score: 500
Sequence: 1 gagctacgcatctgtag.....ccagggcgcgcacacaag 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457241 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_032802:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:*
25: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	748	22	AAD11123
2	500	100.0	1201	22	AAD11122
3	307.8	61.6	1593	23	AA572677
4	302.4	60.5	3093	23	AA572678
5	269.6	53.9	2922	23	AA572679
6	257.6	51.5	518	22	ABA66861
7	257.6	51.5	518	22	ABA63930
8	257.6	51.5	518	22	AAK15297
9	257.6	51.5	518	22	AAK15297

Result No.	Score	Query Match	Length	ID	Description
10	250	50.0	3345	23	AA572676
11	250	50.0	3522	23	AA572321
12	235	47.0	495	22	ABA54299
13	235	47.0	495	22	ABA24056
14	235	47.0	495	22	AAK02583
15	235	47.0	495	22	AAK33949
16	181.4	36.3	1530	21	AA16669
17	178.2	35.6	537	22	AAK22452
18	178.2	35.6	537	22	AAK48620
19	178.2	35.6	537	22	AA124443
20	178.2	35.6	537	22	AA154448
21	178.2	35.6	576	22	ABA50147
22	178.2	35.6	576	22	AA122991
23	178.2	35.6	593	22	AAK09801
24	178.2	35.6	593	22	AAK35695
25	178.2	35.6	593	22	AA141410
26	171	34.2	586	22	ABA60111
27	171	34.2	586	22	AAK08387
28	171	34.2	586	22	AAK34267
29	171	34.2	586	22	AA139989
30	169.2	33.8	468	22	ABA26584
31	169.2	33.8	468	22	AA115206
32	153.6	30.7	471	22	ABA72651
33	153.6	30.7	471	22	AAK21080
34	153.6	30.7	471	22	AAK47234
35	153.6	30.7	471	22	AA153069
36	145.4	29.1	3127	23	AA576376
37	142.4	28.5	471	22	ABA64656
38	142.4	28.5	471	22	AA124385
39	140.8	28.2	459	22	ABA45009
40	140.8	28.2	459	22	AA113771
41	140.8	28.2	526	22	ABA60045
42	140.8	28.2	526	22	AAK08319
43	140.8	28.2	526	22	AAK34197
44	140.8	28.2	526	22	AA139920
45	136	27.2	526	22	ABA61696

ALIGNMENTS

RESULT 1
AAD11123
ID AAD11123 standard; DNA; 748 BP.
XX
AC AAD11123;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human small cell lung cancer associated gene, DKFZp434C196 #2.
XX
KW Human; small cell lung cancer; therapy; hCAP; nucleic acid; NM;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW leiomyosarcoma; synovial sarcoma; cytostatic; ds.
XX
OS Homo sapiens.
XX
PN MO200153349-A2.
XX
PD 26-JUL-2001.
XX
PE 19-JAN-2001; 2001WO-US02015.
XX
PR 21-JAN-2000; 2000US-0489101.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX
DR WPI; 2001-457597/49.
XX

PT Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e.g. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 PS Claim 57; Page 116; 152pp; English.

CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is human DKFZp434C196 DNA. This small cell lung
 CC cancer associated gene is designated as NY-SCLC-13.
 XX

XX Sequence 748 BP; 166 A; 261 C; 246 G; 72 T; 3 other;

Query Match 100.0%; Score 500; DB 22; Length 748;
 Best Local Similarity 100.0%; Pred. No. 1.4e-67;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTCTACGGCATTCGCTGAGAGCGTCCCAAGGCGATCGCTAATGAGAGCGCCGACAGG 60
 DB 1 GAGTCTACGGCATTCGCTGAGAGCGTCCCAAGGCGATCGCTAATGAGAGCGCCGACAGG 60
 QY 61 GCATCGCTAATGAGAGCGACACACAGTGCATCGCCCAAGAGAGCGCCGACAGGCGATCG 120
 DB 61 GCATCGCTAATGAGAGCGACACACAGTGCATCGCCCAAGAGAGCGCCGACAGGCGATCG 120
 QY 121 CCGAGAGCGCATCCGAGGCGATCGCCCAAGAGAGGTTGCCCAAGGCGATCGCCCAATGGGG 180
 DB 121 CCGAGAGCGCATCCGAGGCGATCGCCCAAGAGAGGTTGCCCAAGGCGATCGCCCAATGGGG 180
 QY 121 CCGAGAGCGCATCCGAGGCGATCGCCCAAGAGAGGTTGCCCAAGGCGATCGCCCAATGGGG 180
 DB 121 CCGAGAGCGCATCCGAGGCGATCGCCCAAGAGAGGTTGCCCAAGGCGATCGCCCAATGGGG 180
 QY 181 TCGCCCGCAAGAGGCGATCGCCCAATGAGAGCGCACCCAGGCGATCGCCCAATGGGGCGG 240
 DB 181 TCGCCCGCAAGAGGCGATCGCCCAATGAGAGCGCACCCAGGCGATCGCCCAATGGGGCGG 240
 QY 241 TCCAGGCGCTTCGCAACGGGGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 DB 241 TCCAGGCGCTTCGCAACGGGGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 QY 301 GCATCGCCCAAGAGGCGCGCACCAAGGCGATCGCCCAAGAGAGGTTGCCCAATGGCGATCG 360
 DB 301 GCATCGCCCAAGAGGCGCGCACCAAGGCGATCGCCCAAGAGAGGTTGCCCAATGGCGATCG 360
 QY 361 CTAAAGCGAGCGCGCTTCGCAAGGCGATCGCTAAGAGAGGCGCGCGCGCGCGCGCGCGCG 420
 DB 361 CTAAAGCGAGCGCGCTTCGCAAGGCGATCGCTAAGAGAGGCGCGCGCGCGCGCGCGCGCG 420
 QY 421 AGAGAGCG 480
 DB 421 AGAGAGCG 480
 QY 481 CCGAGGCGATCGCCCAAGG 500
 DB 481 CCGAGGCGATCGCCCAAGG 500

RESULT 2
 AAD11122
 ID AAD11122 standard; DNA; 1201 BP.

AC AAD11122;
 XX
 XX 24-SEP-2001 (first entry)
 DE Human small cell lung cancer associated gene, DKFZp434C196 #1.
 XX

KW Human; small cell lung cancer; therapy: hCAP; nucleic acid; NA;
 KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
 KW leiomyosarcoma; synovial sarcoma; cytostatic; ds.
 XX
 OS Homo sapiens.

XX WO200153349-A2.
 XX
 XX 26-JUL-2001.
 PD
 XX 19-JAN-2001; 2001WO-US02015.
 PF
 XX 21-JAN-2000; 2000US-0489101.
 PR
 XX (LUDWIG) LUDWIG INST CANCER RES.
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
 DR WPI; 2001-457597/49.

XX Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e.g. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 PS Claim 57; Page 114-115; 152pp; English.

CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is human DKFZp434C196 DNA. This small cell lung
 CC cancer associated gene is designated as NY-SCLC-13.
 XX

XX Sequence 1201 BP; 255 A; 377 C; 349 G; 178 T; 42 other;

Query Match 100.0%; Score 500; DB 22; Length 1201;
 Best Local Similarity 100.0%; Pred. No. 1.3e-67;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTCTACGGCATTCGCTGAGAGCGTCCCAAGGCGATCGCTAATGAGAGCGCCGACAGG 60
 DB 1 GAGTCTACGGCATTCGCTGAGAGCGTCCCAAGGCGATCGCTAATGAGAGCGCCGACAGG 60
 QY 61 GCATCGCTAATGAGAGCGACACACAGTGCATCGCCCAAGAGAGCGCCGACAGGCGATCG 120
 DB 61 GCATCGCTAATGAGAGCGACACACAGTGCATCGCCCAAGAGAGCGCCGACAGGCGATCG 120
 QY 121 CCGAGAGCGCATCCGAGGCGATCGCCCAAGAGAGGTTGCCCAAGGCGATCGCCCAATGGGG 180
 DB 121 CCGAGAGCGCATCCGAGGCGATCGCCCAAGAGAGGTTGCCCAAGGCGATCGCCCAATGGGG 180
 QY 181 TCGCCCGCAAGAGGCGATCGCCCAATGAGAGCGCACCCAGGCGATCGCCCAATGGGGCGG 240
 DB 181 TCGCCCGCAAGAGGCGATCGCCCAATGAGAGCGCACCCAGGCGATCGCCCAATGGGGCGG 240
 QY 241 TCCAGGCGCTTCGCAACGGGGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 DB 241 TCCAGGCGCTTCGCAACGGGGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 QY 301 GCATCGCCCAAGAGGCGCGCACCAAGGCGATCGCCCAAGAGAGGTTGCCCAATGGCGATCG 360
 DB 301 GCATCGCCCAAGAGGCGCGCACCAAGGCGATCGCCCAAGAGAGGTTGCCCAATGGCGATCG 360
 QY 361 CTAAAGCGAGCGCGCTTCGCAAGGCGATCGCTAAGAGAGGCGCGCGCGCGCGCGCGCGCG 420

Db 361 CTACACGAGACGCGCCGTCAGGCGATCGCTAACGAGTGGCGCCGCGCAGGCGATCGCCCAACG 420
 QY 421 AGAGACGCGCGCCGAGGATTCGCCAGATGTGCGACAGGCGATCGCCCAACGAGCGCCG 480
 Db 421 AGAGACGCGCGCCGAGGATTCGCCAGATGTGCGACAGGCGATCGCCCAACGAGCGCG 480
 QY 481 CCCAGGCGATCGCCCAACG 500
 Db 481 CCCAGGCGATCGCCCAACG 500

RESULT 3

AAS72677
 ID AAS72677 standard; cDNA, 1593 BP.

AC AAS72677;
 DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8481.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG08490.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 1; SEQ ID No 8481; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS6197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WPIO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 1593 BP; 444 A; 429 C; 434 G; 286 T; 0 other;

Query Match 61.6%; Score 307.8; DB 23; Length 1593;
 Best Local Similarity 78.8%; Pred. No. 1,2e-38;
 Matches 393; Conservative 0; Mismatches 102; Indels 4; Gaps 2;

QY 5 CTACGCGATTGCTGAGAGCGCTCCGAGGCGATGCTATATGAGAGACCGGACGAGGCGAT 64
 Db 873 CGACATCGCTAACGAGGATTCGTCCACGCGATGCTAACGAGGCGCGCGACATAGGCGAT 932
 QY 65 CGTATATGAGGACACACCCAGTCATCGCCCAAGAGAACCGCCGAGGCGATGCG--- 121
 Db 933 CGCCACAGAGAGAGCGCCGACGAGGATCGCCCAAGAGAGAGCGCCGACGATGCGCG 992
 QY 122 CGAGAGCGCCATCGGAGGCGCCCAAGAGAGGTTGCCAGGCGATCGCCATGAGGCT 181
 Db 993 CGAGGATGCGCCGACGAGGCGATCGCCCAAGAGAGTGGCGCCGACGAGGCGATCGCCCAAGAGAG 1052
 QY 182 CGCCGACAGAGGCGATGCGCCATGAGAGAGCGCCAGAGGCGATCGCCCAAGTGGAGCGCGT 241
 Db 1053 CGCGCCGAGGCGATGCGCCACAGAGAGCGCGCCGAGGCGATCGCCCAAGAGAGAGCGCGC 1112
 QY 242 CCACGCGCTTCCGACAGGCGCGCTCTCTCAGCTTCCGACAGGCGAGCGCCGACGG 301
 Db 1113 CCACGCGGATCGCCCAAGAGAGAGCGCGCCGAGGCGATGCGCCCAAGAGAGAGCGCTCACGG 1172
 QY 302 CATCGCCCAAGGAGGAGCGCCCAAGGCGATGCGCCCAAGAGAGTCAACATCGACGCGATCGC 361
 Db 1173 CATCGCCCAAGAGAGAGCGCCCAAGGCGATGCGCCCAAGAGAGAGTGCACGCGATTCGC 1232
 QY 362 TACGAGGAGCGCGCTCAGGCGATGCTAACGAGTGGCGCCGACGAGGCGATCGCCCAAGAG 421
 Db 1233 TACGAGGATGCGCGTCCACGAGGCGATGCTATGAGAGACTCGCTTACGACATGCTGCTATATGA 1292
 QY 422 GAGAGCGCGCCGAGGAGATGCGCCAGAGGCGATGCGCCAGAGGCGCGCGACGAGAGCGCGC 481
 Db 1293 GGTGCGCGTATATGACATGCTATATGACACCGTACAGGCA-GGCTAACGAGAGAGCGCTGT 1351
 QY 482 CCAGGCGATCGCCCAACG 500
 Db 1352 ACACGACATCGCTATATGAG 1370

RESULT 4

AAS72678
 ID AAS72678 standard; cDNA; 3093 BP.

AC AAS72678;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8482.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG08491.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1: SEQ ID No 8482; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 3093 BP; 875 A; 904 C; 862 G; 452 T; 0 other;

Query Match 60.5%; Score 302.4; DB 23; Length 3093;
Best Local Similarity 79.8%; Pred. No. 7, 2e-38;
Matches 389; Conservative 0; Mismatches 91; Indels 6; Gaps 2;

QY 24 GCTGCCAAGGAGCTTCCTATAGAGACCCGACCAAGGCGATCGTATATGAGACACC 83
DB 1990 GATGCCAAGAACTCCCTATGAGAGACGACGACGATCGTATGAGAGAGGCGACA 2049
QY 84 CAGTGCATGCCCAAG 140
DB 2050 CAGAGCAATGCTCTAG 2109
QY 141 ATGCCCAAG 200
DB 2110 ATGCCCAAG 2169
QY 201 AATGAG 260
DB 2170 AGCAG 2229
QY 261 GAGGCGCTGCTGAGCTTGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 320
DB 2230 GAGGCGCTGAGCTTGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2289
QY 321 ACCAAGGAG 380
DB 2290 GCGGAG 2349
QY 381 GCGATGCTTACAG 440
DB 2350 GCGATGCTTACAG 2409
QY 441 GC---CGAGAGATGTCAG 497
DB 2410 GCGAAG 2469

RESULT 5
ID AAS72679 standard; cDNA; 2922 BP.
XX AAS72679;
AC AAS72679;

XX 13-FEB-2002 (first entry)
DT DNA encoding novel human diagnostic protein #8483.
DE Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HSE-) HXSEQ INC.
XX Dymnac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG08492.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1: SEQ ID No 8483; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 2922 BP; 833 A; 829 C; 806 G; 454 T; 0 other;

Query Match 53.9%; Score 269.6; DB 23; Length 2922;
Best Local Similarity 75.4%; Pred. No. 6, 3e-33;
Matches 362; Conservative 0; Mismatches 114; Indels 4; Gaps 2;

QY 24 GCTGCCAAGGAGCTTCCTATAG 83
DB 1906 GATGCCAAGAAAGTCCCTATGAG 1965
QY 84 CAGTGCATGCCCAAG 140
DB 1966 CAGTGCATGCCCAAG 2025
QY 141 ATGCCCAAG 200
DB 2025 ATGCCCAAG 2085

Qy	201	AAAGAGAGCGCACCACCGAGGCATCGCCCAACTGGAGCGCGCTGCACGGCTTCGCAACGGG	260
Db	2086	AGGAGAGAGCGCGCCCGCAGGCGATCGCCCAACGAGGTCGCGCGCCGAGGCGATCGCAACGAG	2145
Qy	261	GAGCGGTCCTACGTTTCGCCAAGGGAGAGCGCGCCCGAGGCGATCGCCCAACGGGAGCGC	320
Db	2146	GAGCGCGCCCGCAGGCGATTCGCCCAAGAGAGAGCGCTTCAGGCGCATTCGCCCAAGAGAGCGCC	2205
Qy	321	ACCAAGGCGATGGGCAGACGAGGTCACCATTCGCGCCGCGCATCGTTACGAGAGACGCGTTCGAG	380
Db	2206	GCCCAAGCGATTCGCCCAACGAGAGAGCGTTCGCCCAAGGCGCATTCGTTACGAGAGATGCGGTGCAC	2265
Qy	381	GCGATTCGTCACGAGGTCGGCGCGCCCGAGGCGATTCGCCCAAGAGAGAGCGCGCCCGCGGAGATC	440
Db	2266	GCGATTCGTCATGAGAGACTCCGTCATACGACATTCGTCATATGAGATGCGCATATATGACATC	2322
Qy	441	GCGGAGAGTTCGCAAGGCGATTCGCCCAAGAGAGAGCGCGCCCGAGGCGATTCGCCCAACAG	500
Db	2326	GCTATGACACCGTCACAGGCA-GCGTTAAGAGAGAGCGCTGTACACGACATTCGTCATATGAG	2388

RESULT 6
ADP466961

ID	ABA66861	standard; DNA; 518 BP.
XX	ABA66861;	
AC		
XX		
DT	01-FEB-2002	(first entry)
XX		
DE	Human foetal liver single exon nucleic acid probe #15166.	
XX		
KN	Human; foetal liver; gene expression; single exon nucleic acid probe; ss	

XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0652366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 15166; 639bp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at http://wipo.int/pub/publiced_pct_sequences.
XX Sequence 518 BP; 140 A; 170 C; 139 G; 69 T; 0 other;

Best Local Similarity 75.2%; Pred No. 5.1e-31;
Matches 376; Conservative 0; Mismatches 114; Indels 10; Gaps 4;

OY	5	CTACGGCATTTGCTGAGAGACGCTTCCCAAGGGGCATTCGGTAAATGAGAGAGCGGCACAGGGGCAT	6
Db	23	CAACATCACTAACAGAGAGATGCCCTACACGATATCACTAACAGAGAGCGCTGTCCACGGGCAT	82
OY	65	CGCTATATGAGACACACACCGTGCATTCGGCCAAAGAGAAAGCGGCCCAAGGCATATCGC---	121
Db	83	CGCTAACAGAGCGCGCGGCCCAAGGGGCATTCGGCCAAAGAGAGAGCGGCCCAAGGGGCATTCGCCAA	142
OY	122	CGAGAGCGCCATCCAGAGGCATTCGCCAAAGAGAGAGGTTGCCCAAGGGCATTCGGCATGGGGT	181
Db	143	CGAGAGCGCGCCCGCCAGGAATTCGCACAGAGAGAGAGCGCGGCCCAAGGGGCATTCGGCCAAAGAGT	202
OY	182	CGCGGCACAGGGGCATTCGGCAATTAAGAGAGCGCCACCGAGGGCATTCGGCCAAATCGGAGCGCGT	241
Db	203	CGCGCGCCCAAGGGGCATTCGGCCAAAGAGAGAGCGCGGCCCATTCGGCCAAAGAGAGCGCGCT	262
OY	242	CGACGGCTTCGGCCAAAGGGAGCGCGGCATTCACATTCGGCCAAAGGGAGCGCGGCCCAAGG	301
Db	263	CGATGGCATTCGGCCAAAGAGAGAGCGTGCCTCAAGGCATTCGTTAACAGAGAGATGCGCGTACAGG	322
OY	302	CATTCGCCAAAGGGAGAGCGCCACCAAGAGGCATTCGGCCAAAGAGAGTCCATTCACAGGGCATTCG	361
Db	323	CATTCGGTAAATGAGAGCTCGTATACGACATTCGGCTAAATGAGAGTGCATTAATATGACATTCGC	382
OY	362	TAAAGAGAGCGCGTCCAGGGGCATTCGGTTAAAGAGAGTGGCGCGGCCCAAGGCATTCGCCAACGA	421
Db	383	TA---ATGACACCGTAAACAAGGCA--CGCTAACAGAGAGATTCGTTAAACAGCATATTCGGTTATGA	438
OY	422	GGAAGCGCGGCCAGGGAATTCG---CCGAGGATATTCGCAACAGAGGCATTCGCCAAAGAGAGACG	478
Db	439	GGAACCATTAACAAGGCATTCGGTAAATGAGAGCGTTGTATATGACATTCGCTAACAAAGAGAAC	498
OY	479	CGGCCAGAGGCATTCGCCAACGA	498
Db	499	TCCTAACAGCGCTCGCTAACGA	518

RESULT 7

PA	ABA33930 standard; DNA; 518 BP.
XX	
AC	ABA33930;
XX	
DT	23-JAN-2002 (first entry)
XX	
DE	Probe #12396 for gene expression analysis in human heart cell sample
XX	
XX	Human; gene expression; heart; microarray; vascular system; probe;
KM	cardiovascular disease; hypertension; cardiac arrhythmia;
KM	congenital heart disease; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200157274-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-0500666.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000SB-0024263.
XX	
XX	

(MOLE-) MOLECULAR DYNAMICS INC.

XX WPI: 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -

PS Claim 4: SEQ ID NO 12396; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 518 BP; 140 A; 170 C; 139 G; 69 T; 0 other;

Query Match 51.5%; Score 257.6; DB 22; Length 518;
Best Local Similarity 75.2%; Pred. No. 5.1e-31;
Matches 376; Conservative 0; Mismatches 114; Indels 10; Gaps 4;

QY 5 CTACGGCATTTGCTGAGAGCGCTGCCAGGCGATCGCTAATGAGAGCGCCAGGCGCAT 64
DB 23 CAACATCTACTAACGAGATGCTCTACAGATATCTACAGAGAGCGCTCTCCAGCGCAT 82
QY 65 CGCTAATGAGAGACACCCAGTGCATCGCCACAGAGAGAGCGCCAGGCGCATCGC--- 121
DB 83 CGCTAATGAGAGAGCGCGCGCCAGGCGATCGCCACAGAGAGAGCGCCAGGCGCATCGCAA 142
QY 122 CGAGAGCGCATTCAGAGGCGATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 181
DB 143 CGAGAGCGCGCGCGCCAGGCGATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 202
QY 182 CGCGGCGAGAGCGCGCGCCAGGCGATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 241
DB 203 CGCGGCGCGCGCGCGCCAGGCGATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 262
QY 242 CGAGCGCTTCGCGCAACGGGAGCGCGCTCGTTCGCGCAAGGAGAGCGCGCGCGCG 301
DB 263 CGATGCGATCGCGCAACGGGAGCGCGCTCGTTCGCGCAAGGAGAGCGCGCGCGCG 322
QY 302 CATCGCGAGAGAGCG 361
DB 323 CATCGCGATGAGAGACTCGCTATAGAGATCGCTAATGAGAGAGCGCGCGCGCGCG 382
QY 362 TAACGAGAGCG 421
DB 383 TA---ATGACACCGTACAGAGGCA-CGCTTACGAGAGAGTGTACACAGCTCGCTATATA 438
QY 422 GAGAGCG 478
DB 439 GAGAGCG 498
QY 479 CGCGCGAGGCGCATCGCGCAACA 498
DB 499 TCTACAGAGCGCGTCTACACA 518

RESULT 8
AAK15297 standard; DNA; 518 BP.

XX AAK15297;
AC AAK15297;
XX 05-NOV-2001 (first entry)
DT Human brain expressed single exon probe SEQ ID NO: 15288.
XX DE

XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-0500667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207436.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4: SEQ ID NO: 15288; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is one of the probes of the

XX invention.

XX Sequence 518 BP; 140 A; 170 C; 139 G; 69 T; 0 other;

XX Query Match 51.5%; Score 257.6; DB 22; Length 518;

XX Best Local Similarity 75.2%; Pred. No. 5.1e-31;

XX Matches 376; Conservative 0; Mismatches 114; Indels 10; Gaps 4;

QY 5 CTACGGCATTTGCTGAGAGCGCTGCCAGGCGATCGCTAATGAGAGCGCCAGGCGCAT 64
DB 23 CAACATCTACTAACGAGATGCTCTACAGATATCTACAGAGAGCGCTCTCCAGCGCAT 82
QY 65 CGCTAATGAGAGACACCCAGTGCATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 121
DB 83 CGCTAATGAGAGAGCGCGCGCCAGGCGATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 142
QY 122 CGAGAGCGCATTCAGAGGCGATCGCCACAGAGAGAGTGGTCCAGGCGCATGAGG 181
DB 143 CGAGAGCG 202
QY 182 CGCGGCGAGAGCG 241
DB 203 CGCGGCG 262
QY 242 CGAGCGCTTCGCGCAACGGGAGCGCGCTCGTTCGCGCAAGGAGAGCGCGCGCGCG 301
DB 263 CGATGCGATCGCGCAACGGGAGCGCGCTCGTTCGCGCAAGGAGAGCGCGCGCGCG 322
QY 302 CATCGCGAGAGAGCG 361
DB 323 CATCGCGATGAGAGACTCGCTATAGAGATCGCTAATGAGAGAGCGCGCGCGCGCG 382
QY 362 TAACGAGAGCG 421
DB 383 TA---ATGACACCGTACAGAGGCA-CGCTTACGAGAGAGTGTACACAGCTCGCTATATA 438

QY 422 GGACGCCGCCGAGGATG---CCGAGATGTCGACAGGCGATGCCAAGAGAGCC 478
 DB 439 GGACGCCGCCGAGGATG---CCGAGATGTCGACAGGCGATGCCAAGAGAGCC 498
 QY 479 CGCCGAGGCGATGCCAACA 498
 DB 499 TCTACAGCGCTGCTTACCA 518

RESULT 9

AA147072
 ID AA147072 standard; DNA; 518 BP.

AA147072:

17-OCT-2001 (first entry)

DE Probe #15758 used to measure gene expression in human placenta sample.

KM Probe: microarray; human; placenta; antenatal diagnosis;
 genetic disorder; ss.

OS Homo sapiens.

PN W0200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000US-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 15758; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders.

XX Sequence 518 BP; 140 A; 170 C; 139 G; 69 T; 0 other;

QY Query Match 51.5%; Score 257.6; DB 22; Length 518;

DB Best Local Similarity 75.2%; Pred. No. 5.1e-31;

Matches 376; Conservative 0; Mismatches 114; Indels 10; Gaps 4;

QY 5 CTAGGGGATTGCTGAGGAGCTGCGGAGGCGCTAATGATGAGCGCCGACGAGGCGAT 64

DB 23 CAACTATCCTAAGAGGATGCTGCTACAGATATCTACTAGAGAGCGCTGTCCAGGCGAT 82

QY 65 CGCTAATGAGGAGACCAACCAAGTGCATGCCCAAGAGAGAGCGCCGAGGCGATGCG--- 121

DB 83 CGCTAAGAGGCGCGCGCCGAGGCGATGCCCAAGAGAGAGCGCCGAGGCGATGCGCAA 142

QY 122 CGAGAGCGGATCCAGGCGATGCCCAAGAGAGAGGTTGCCGAGGCGATGCCAATGGGCT 181

DB 143 CGAGAGCGGCGCGCGCCGAGGCGATGCCCAAGAGAGAGCGCCGAGGCGATGCCCAAGAGGCT 202

QY 182 CGCGCCAGAGGCGATGCCCAATGAGAGAGCGCCAGCGGCGATGCCCAAGAGAGCGCGCT 241
 DB 203 CGCGCCAGAGGCGATGCCCAATGAGAGAGCGCGCGCGGCGATGCCCAAGAGAGAGCGCGCT 262
 QY 242 CCAGGCGCTTGGCCAGAGGAGAGCGCGCTCTGAGCTTGGCCAGAGGAGAGCGCGCGCGAGG 301
 DB 263 CCAGGCGATGCCCAAGAGAGAGCGCTGCGCGGCGATGCCCAAGAGAGAGCGCGCGCGAGG 322
 QY 302 CATTGCCCAAGAGGAGAGCGCGCGCGGCGATGCCCAAGAGAGAGCGCGCGCGAGGCGCGCG 361
 DB 323 CATTGCCCAATGAGAGAGAGCGCGCGCGGCGATGCCCAATGAGAGAGAGCGCGCGCGAGGCGCG 382
 QY 362 TAAAGAGAGAGCGCGCGCGGCGATGCCCAAGAGAGAGGCGCGCGCGGCGATGCCCAAGAG 421
 DB 383 TA---ATGACAGCGGTACAGAGCA -CGTAAAGAGAGAGGCTGTACAGAGAGCGCGTAAATGA 438
 QY 422 GGACGCCGCCGAGGATG---CCGAGATGTCGACAGGCGATGCCAAGAGAGAGCC 478
 DB 439 GGACGCCGCCGAGGATG---CCGAGATGTCGACAGGCGATGCCAAGAGAGAGCC 498
 QY 479 CGCCGAGGCGATGCCAACA 498
 DB 499 TCTACAGCGCTGCTTACCA 518

RESULT 10

AA572676
 ID AA572676 standard; cDNA; 3345 BP.

AA572676:

13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8480.

KM Human: chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR P-PSDB; ABG08489.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity -

PS Claim 1; SEQ ID No 8480; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC for restoring normal activity of (II) or to treat disease states involving

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://wipo.int/pub/published_pct_sequences.

XX Sequence 3345 BP; 954 A; 940 C; 913 G; 538 T; 0 other;

Query Match 50.0%; Score 250; DB 23; Length 3345;
 Best Local Similarity 74.1%; Pred. No. 5,6e-30;

Matches 372; Conservative 0; Mismatches 120; Indels 10; Gaps 4;

OY 5 CTACGGCATTCGTGAGGAGCGCTGCCAGGCGATCGCTAATGAGAGCGCCGACCGGCGAT 64
 DB 2076 CAACATCTACTACGAGGATGCTTACACGATATCTACTACGAGAGCGCTGTCCAGCGCT 2135
 OY 65 CGCTAATGAGAGACACCCAGCTGATCCGACAGAGAAAGCCGCCAGGCGATCGC--- 121
 DB 2136 CGCTAATGAGAGCGCGCCGACAAAGGATGCGCCACAGAGAGAGCGCCAGGCGATCGCAA 2195
 OY 122 CGAGAGCGCCATCCAGGCGATCGCCACAGAGAGGTTGCCAGGCGATCGCCATATGGGAT 181
 DB 2196 CGAGAGCGCGCGCCGACGAGATGCGCCAGGAGAGCGCCGCGCCAGGCGATCGCCACAGCGAT 2255
 OY 182 CGCCGACAGCGGCGATCGCCATGAGAGAGCGCCAGGCGATCGCCATGAGGAGCGCGT 241
 DB 2256 CGCCGCGCCAGGCGGATCGCCACAGAGAGCGCGCCAGGCGATCGCCACAGAGAGCGCGC 2315
 OY 242 CGAGGCGTTCGCGCAAGGAGAGCGCGCTCTGCTTACGCGCAAGGAGAGCGCGCGCG 301
 DB 2316 CGAGGCGATTCGCGCAAGGAGAGGATGCTGCCAGGCGATTCGTAATGAGAGAGGCGGTGACGG 2375
 OY 302 CATCGCGCAAGGAGAGCGCGCCACCAAGGCGATGAGAGAGCGCGATCGCCATCGCG 361
 DB 2376 CATCGGTATGAGAGACTCGCTATACGATCGCTAATGAGAGAGCGCGATTCGATACATCGC 2435
 OY 362 TAACGAGAGAGCGCGCTCCAGGCGATCGCTAATGAGAGAGTGGCCGCGCGCGATCGCCACG 421
 DB 2436 TA---ATGACACCGTACAGAGGCA-CACTAAGAGAGAGCGCTGTACAGACATCGCTAATGA 2491
 OY 422 GGAAGCGCGCCAGGAGATCG---CCGAGAGATGTCGACAGAGGCGATCGCCACAGAGAGCG 478
 DB 2492 GGAAGCGCGATTCAGAGGATGCTGTAATGAGAGAGCGTGTATATGACATCGCTAATGAGAG 2551
 OY 479 CGCCGAGGCGATCGCCACAG 500
 DB 2552 TCTACAGAGCGCTGCTACACAG 2573

RESULT 11

AAS72321 ID AAS72321 standard; cDNA; 3522 BP.

XX AAS72321;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8125.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-0508631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG08134.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 1; SEQ ID No 8125; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC quantifying a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations in

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp://wipo.int/pub/published_pct_sequences.

XX Sequence 3522 BP; 1005 A; 1006 C; 959 G; 553 T; 0 other;

Query Match 50.0%; Score 250; DB 23; Length 3522;

Best Local Similarity 74.1%; Pred. No. 5,6e-30;

Matches 372; Conservative 0; Mismatches 120; Indels 10; Gaps 4;

OY 5 CTACGGCATTCGTGAGGAGCGCTGCCAGGCGATCGCTAATGAGAGCGCCGACCGGCGAT 64
 DB 2253 CAACATCTACTACGAGGATGCTTACACGATATCTACTACGAGAGCGCTGTCCAGCGCT 2135
 OY 65 CGCTAATGAGAGACACCCAGCTGATCCGACAGAGAAAGCCGCCAGGCGATCGC--- 121
 DB 2136 CGCTAATGAGAGCGCGCCGACAAAGGATGCGCCACAGAGAGAGCGCCAGGCGATCGCAA 2195
 OY 122 CGAGAGCGCCATCCAGGCGATCGCCACAGAGAGGTTGCCAGGCGATCGCCATATGGGAT 181
 DB 2196 CGAGAGCGCGCGCCGACGAGATGCGCCAGGAGAGCGCCGCGCCAGGCGATCGCCACAGCGAT 2255
 OY 182 CGCCGACAGCGGCGATCGCCATGAGAGAGCGCCAGGCGATCGCCATGAGGAGCGCGT 241
 DB 2256 CGCCGCGCCAGGCGGATCGCCACAGAGAGCGCGCCAGGCGATCGCCACAGAGAGCGCGC 2315
 OY 242 CGAGGCGTTCGCGCAAGGAGAGCGCGCTCTGCTTACGCGCAAGGAGAGCGCGCG 301
 DB 2316 CGAGGCGATTCGCGCAAGGAGAGGATGCTGCCAGGCGATTCGTAATGAGAGAGGCGGTGACGG 2375
 OY 302 CATCGCGCAAGGAGAGCGCGCCACCAAGGCGATGAGAGAGCGCGATCGCCATCGCG 361
 DB 2376 CATCGGTATGAGAGACTCGCTATACGATCGCTAATGAGAGAGCGCGATTCGATACATCGC 2435
 OY 362 TAACGAGAGAGCGCGCTCCAGGCGATCGCTAATGAGAGAGTGGCCGCGCGCGATCGCCACG 421
 DB 2436 TA---ATGACACCGTACAGAGGCA-CACTAAGAGAGAGCGCTGTACAGACATCGCTAATGA 2491
 OY 422 GGAAGCGCGCCAGGAGATCG---CCGAGAGATGTCGACAGAGGCGATCGCCACAGAGAGCG 478
 DB 2492 GGAAGCGCGATTCAGAGGATGCTGTAATGAGAGAGCGTGTATATGACATCGCTAATGAGAG 2551
 OY 479 CGCCGAGGCGATCGCCACAG 500
 DB 2552 TCTACAGAGCGCTGCTACACAG 2573

QY 362 TAACGAGACGCCGTCCAGGCGATCGTACGAGAGTGGCCGCCAGGCGCATCGCCACGA 421
 DB 2613 TA---ATGACACCGTGTACAGAGCA-CAGTAAAGCAGAGCGCTGTACAGCATGCTTAATGA 2668
 QY 422 GGAGCGCCGCCAGGGAATG---CCGAGATGTGCGACAGGCGATCGCCACAGAGAGCGC 478
 DB 2669 GGACACCATACAGAGCATGTGTAATGAGAGCTGTATATGACATCGCTAACAGAGACAC 2728
 QY 479 CGCCAGGCGCATCGCCACAG 500
 DB 2729 TCTACAGCGCTCGCTAACAG 2750

RESULT 12

ABR54299
 ID ABR54299 standard; DNA; 495 BP.

AC ABR54299;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #2604.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO20015727-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human foetal liver -

XX Claim 1; SEQ ID NO 2604; 639bp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 495 BP; 121 A; 174 C; 129 G; 71 T; 0 other;

XX Query Match 47.0%; Score 235; DB 22; Length 495;

XX Best local similarity 77.3%; Pred. No. 1.3e-27;

XX Matches 299; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 43 ATGAGAGAGCGCGCATGCTTAATGAGACAGCAGCCATGCGTCCAGAGAGG 102

DB 108 ATGAGAGAGCGCGCATGCTTAATGAGACAGCAGCCATGCGTCCAGAGAGG 167

QY 103 AAGCGCGCGCATGCG---CGAGAGCGCATGCGAGGCGCGCCAGAGAGAGGTTG 159

DB 168 AGGTGTTCACGAGCATGCTTAATGAGAGCGCGCGCCAGGCGCATCGCCACAGAGAGCGCG 227
 QY 160 CCCAGGCGCATCGCCATGAGGCGCGCGCCAGCAGGCGCATCGCCATGAGAGAGCGCCACAGG 219
 DB 228 CCCAGGCGCATCGCCACAGAGAGCGCGCGCCAGGAAATGCGACAGGAGCGCGCGCGAGG 287
 QY 220 GCATCGCAACTGCGAGCGCGCTTCAGCAGGCGCTTCAGCAGGAGCGCGCTTCAGCTTCG 279
 DB 288 GCATCGCGCAACAGAGGTCGCGCGCCAGGCGCATCGCCACAGAGAGCGCGCGCCAGGCGATG 347
 QY 280 CCAGCGGAGAGCGCGCGCGCGCGCCAGGCGCATCGCCACAGGAGCGCGCGCGCGCGCAAG 339
 DB 348 CCAGCGGAGAGCGCGCGCTTCAGGCGCATCGCCACAGAGAGCGCGCGCGCGCATTCGTAAG 407
 QY 340 AGGTGACATTCAGCGCGCTTAATGAGAGCGCGCTTCAGGCGCATCGCCATGAGAGGTCG 399
 DB 408 AGGATGCGGTGCGAGCGCGCTTAATGAGAGCGCGCTTCAGGCGCATCGCCATGAGAGATG 467
 QY 400 CCGCGCGAGGCGATCGCCACAGAGAGCG 426
 DB 468 CCATATATGACATCGCTTAATGAGACCG 494

RESULT 13

ABR24056
 ID ABR24056 standard; DNA; 495 BP.

AC ABR24056;

DT 23-JAN-2002 (first entry)

DE Probe #2522 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

XX Homo sapiens.

XX WO20015727-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

XX Claim 1; SEQ ID NO 2522; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart. The

XX present sequence is one such probe. The probes may be used for

XX predicting, measuring and displaying gene expression in samples derived

XX from the human heart via microarrays. By measuring gene expression, the

XX probes are useful for predicting, diagnosing, grading, staging,

XX monitoring and prognosing diseases of the human heart and vascular system

XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 495 BP; 121 A; 174 C; 129 G; 71 T; 0 other;
 SQ
 Query Match 47.0%; Score 235; DB 22; Length 495;
 Best Local Similarity 77.3%; Pred. No. 1.3e-27;
 Matches 299; Conservative 0; Mismatches 85; Indels 3; Gaps 1;
 XX
 YY 43 ATGAGGACGCGGACGCGGATGCTATGAGACACACCGGATGCTGCGCAAGAGG 102
 DB 108 ATGAGGACGCGGACGCGGATGCTATGAGACACACCGGATGCTGCGCAAGAGG 167
 YY 103 AAGCGCGCCAGGCGATGCG---CGAGGACGCGGATGCTGCGCAAGAGGAGTTG 159
 DB 168 ACGGTGTCACGCGGATGCTATGAGACGCGGCGCGGCGGCGGCGGCGGCGGCGG 227
 YY 160 CCGAGGCGATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 219
 DB 228 CCGAGGCGATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 287
 YY 220 GCATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 279
 DB 288 GCATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 347
 YY 280 CCAACGGGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 339
 DB 348 CCAACGGGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 407
 YY 340 AGGTACACATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 399
 DB 408 AGGTACACATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 467
 YY 400 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
 DB 468 CCATATATGACATGCTATGACACCG 494
 RESULT 14
 ID AAK02583 standard; DNA; 495 BP.
 AC AAK02583;
 XX
 YY 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe SEQ ID NO: 2574.
 KW Human: brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX
 OS Homo sapiens.
 PN WO200157275-A2.
 PD 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 YY 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI

XX WPI; 2001-483446/52.
 DR
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4: SEQ ID NO: 2574; 650bp + Sequence Listing; English.
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX
 SQ Sequence 495 BP; 121 A; 174 C; 129 G; 71 T; 0 other;
 XX
 Query Match 47.0%; Score 235; DB 22; Length 495;
 Best Local Similarity 77.3%; Pred. No. 1.3e-27;
 Matches 299; Conservative 0; Mismatches 85; Indels 3; Gaps 1;
 XX
 YY 43 ATGAGGACGCGGACGCGGATGCTATGAGACACACCGGATGCTGCGCAAGAGG 102
 DB 108 ATGAGGACGCGGACGCGGATGCTATGAGACACACCGGATGCTGCGCAAGAGG 167
 YY 103 AAGCGCGCCAGGCGATGCG---CGAGGACGCGGATGCTGCGCAAGAGGAGTTG 159
 DB 168 ACGGTGTCACGCGGATGCTATGAGACGCGGCGCGGCGGCGGCGGCGGCGGCGG 227
 YY 160 CCGAGGCGATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 219
 DB 228 CCGAGGCGATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 287
 YY 220 GCATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 279
 DB 288 GCATGCGCAATGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 347
 YY 280 CCAACGGGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 339
 DB 348 CCAACGGGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 407
 YY 340 AGGTACACATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 399
 DB 408 AGGTACACATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 467
 YY 400 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
 DB 468 CCATATATGACATGCTATGACACCG 494
 RESULT 15
 ID AAI33949 standard; DNA; 495 BP.
 AC AAI33949;
 XX
 YY 17-OCT-2001 (first entry)
 DE Probe #2635 used to measure gene expression in human placenta sample.
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200157272-A2.
 PD 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00663.
 PR 04-FEB-2000; 2000US-0180312.
 XX

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI; 2001-488897/53.

XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -

XX
PS Claim 25; SEQ ID No 2635; 654pp; English.

XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX
SQ Sequence 495 BP; 121 A; 174 C; 129 G; 71 T; 0 other;

Query Match 47.0%; Score 235; DB 22; Length 495;

Best Local Similarity 77.3%; Pred. No. 1,3e-27;
Matches 299; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 43 ATGAGGACCCGACGCGGATCGCTAATGAGACACCCAGTGCATGCCACAGAG 102
DB 108 ATGAGGACCCGATACACACATCTACTACAGAGAGATGCTCTACAGATATCTAACGAGG 167
QY 103 AAGCCGCCAGGCGATCGC---CGAGAGCGCCATCCAGGCGATGCCAACGAGAGGTTG 159
DB 168 AGCGTGTCCAGGCGATCGCTAACGAGGCGCGCCGCGGCGATGCCAACGAGAGCGCG 227
QY 160 CCCAGGCGATCGCCATGAGGCGCGCGCACAGAGGCGATGCCAATGAGAGCGCCACCCAG 219
DB 228 CCCAGGCGATCGCCACAGGAGGAGCGCGCCGCGGATGCCAGGAGAGCGCGCCAG 287
QY 220 GCATCGCACTGAGAGCGCGTTCACAGGCGTTCGCAAGGAGGAGCGCGTCTGAGCTTG 279
DB 288 GCATCGCCAGAGAGGCGCGCGCCGCGGCGGCGATGCCAAGAGAGCGCGCCAGGCGATCG 347
QY 280 CCAAGGAGGAGCGCGCGCGCGGCGATGCCAAGGAGGAGCGCCACCAAGGCGATGCCAAGC 339
DB 348 CCAAGGAGGAGCGCGCGTTCATGCGATGCCAAGGAGAGCGCGTCCAGGCGATGCCAAGC 407
QY 340 AGGTACGATCCAGGCGATGCCAAGGAGAGCGCGTTCAGGCGATGCCAAGGAGAGTGG 399
DB 408 AGGATCGCGTACGAGCGATGCCAAGGAGAGCGCGTTCAGGCGATGCCAAGGAGAGTGG 467
QY 400 CCGCCCGAGGCGATGCCAAGGAGAGCG 426
DB 468 CCATATATGATCGCTAATGAGACCG 494

Search completed: October 10, 2002, 15:18:05
Job time : 93 secs

Result	Score	Match	Query	length	DB	ID	Description
1	116.4	23.3	390	4	US-09-197-649-7	Sequence 7, Appl1	
c 2	87.4	17.5	4403765	4	US-09-103-8640A-2	Sequence 2, Appl1	
c 3	82.8	16.6	398	4	US-09-197-649-7	Sequence 7, Appl1	
c 4	77.2	15.4	1140	3	US-09-023-173-4	Sequence 4, Appl1	
5	72.7	15.4	4494	4	US-09-056-556-176	Sequence 176, Appl1	
c 6	76.8	12.4	1931	2	US-09-130-114-2	Sequence 2, Appl1	
c 7	76.2	13.2	1439	4	US-09-056-556-167	Sequence 167, Appl1	
8	74.4	14.9	1248	4	US-09-105-531-7	Sequence 7, Appl1	
c 9	74.4	14.9	1439	4	US-09-056-556-167	Sequence 167, Appl1	
c 10	74.4	14.9	1613	4	US-09-105-531-3	Sequence 3, Appl1	
c 11	74	14.8	5036	4	US-09-177-343-2	Sequence 2, Appl1	
c 12	74	14.8	5970	3	US-09-320-878-21	Sequence 21, Appl1	
13	73.4	14.7	12588	2	US-08-387-942C-1	Sequence 1, Appl1	
14	73.2	14.6	15872	4	US-09-105-537-1	Sequence 1, Appl1	
15	72.6	14.5	1294	3	US-09-025-691-92	Sequence 2, Appl1	
16	72.4	14.5	1182	4	US-09-385-028-19	Sequence 19, Appl1	
17	72.4	14.5	11604	4	US-09-385-028-13	Sequence 13, Appl1	
18	72.4	14.5	15079	4	US-09-385-028-11	Sequence 11, Appl1	
19	72.2	14.4	400	4	US-09-056-556-179	Sequence 179, Appl1	
20	72.2	14.4	3306	4	US-08-261-206A-71	Sequence 71, Appl1	
c 21	72	14.4	494	4	US-09-056-556-116	Sequence 116, Appl1	
c 22	72	14.4	985	4	US-09-056-556-182	Sequence 182, Appl1	
c 23	71.8	14.4	778	3	US-08-356-307B-11	Sequence 11, Appl1	
c 24	71.8	14.4	778	3	US-08-356-307B-11	Sequence 11, Appl1	
c 25	71.2	14.2	4411528	4	US-09-103-8640A-1	Sequence 1, Appl1	
c 26	70.8	14.2	2712	4	US-09-025-691-4	Sequence 4, Appl1	
27	70.8	14.2	4403765	4	US-09-103-8640A-2	Sequence 2, Appl1	

28	70.8	14.2	4411539	4	US-09-123-840A-1	Sequence 1, Appl
29	70.2	14.0	30001	4	US-08-105-553-1	Sequence 1, Appl
30	70.2	14.0	30001	4	US-08-474-973-1	Sequence 1, Appl
31	69.2	13.8	985	4	US-09-056-556-182	Sequence 182, App
32	69	13.8	12588	2	US-08-387-942C-1	Sequence 1, Appl
33	67.6	13.5	1193	4	US-09-372-422A-23	Sequence 23, Appl
34	67	13.4	281	2	US-08-419-075-4	Sequence 4, Appl
35	66.8	13.4	352	2	US-08-419-075-25	Sequence 25, Appl
36	66.2	13.2	1187	1	US-08-440-856A-2	Sequence 2, Appl
37	66	13.2	579	2	US-08-737-524B-17	Sequence 17, Appl
38	66	13.2	579	2	US-08-419-075-3	Sequence 3, Appl
39	66	13.2	639	2	US-08-737-524B-12	Sequence 12, Appl
40	66	13.2	639	2	US-08-419-075-5	Sequence 2, Appl
41	66	13.2	647	2	US-08-419-075-20	Sequence 20, Appl
42	66	13.2	653	2	US-08-419-075-18	Sequence 18, Appl
43	66	13.2	2123	2	US-08-737-524B-9	Sequence 9, Appl
44	66	13.2	2123	2	US-08-419-075-1	Sequence 1, Appl
45	65.8	13.2	439	2	US-08-387-942C-31	Sequence 31, Appl

RESULT 1
US-09-197-649-7

```

1  Patent No. 6194550
2  GENERAL INFORMATION:
3  APPLICANT: Gold, Larry
4  APPLICANT: Tuerk, Craig
5  APPLICANT: Pribnow, David
6  APPLICANT: Smith, Jonathan D.
7  TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
8  FILE REFERENCE: NEX02/C1-CON
9  CURRENT APPLICATION NUMBER: US/09/197,649
10 CURRENT FILING DATE: 1998-11-23
11 EARLIER APPLICATION NUMBER: 07/829,461
12 EARLIER FILING DATE: 1992-01-31
13 EARLIER APPLICATION NUMBER: 07/739,055
14 EARLIER FILING DATE: 1991-08-01
15 EARLIER APPLICATION NUMBER: 07/561,968
16 EARLIER FILING DATE: 1990-08-02
17 NUMBER OF SEQ ID NOS: 26
18 SOFTWARE: PatentIn Ver. 2.0
19 SEQ ID NO 7
20 LENGTH: 390
21 TYPE: DNA
22 ORGANISM: Artificial Sequence
23 FEATURE:
24 OTHER INFORMATION: Description of Artificial Sequence: Sequence
25 OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
26 OTHER INFORMATION: fragments having NcoI restriction sites.
27 JS-09-197-643-7

```

[illegible]

Fri Oct 11 09:30:45 2002

us-09-489-101a-15_copy_1_500.rni

Page 2

```
OY 353 CGGATCGCTTAACAGAGACCGGTCAGGCGCATCGCTAAGAGAGTGGCGCCAGGGGAT 412
    || || || || || || || || || || || || || || || || || || || || ||
Db 242 CGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
OY 413 CGCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472
    || || || || || || || || || || || || || || || || || || || || ||
Db 302 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
OY 473 GGAAGCGCGCCAGAG 486
    || || || || || || || || || || || || || || || || || || || || ||
Db 362 CGAGAGAGAGAGAG 375
```

```
RESULT 2
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 17.5%; Score 87.4; DB 4; Length 4403765;
Best Local Similarity 49.2%; Pred. No. 6.5e-08;
Matches 224; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

OY 33 GGCATCGCTAATGAGAGACCGCGACCGAGGCGATCGCTAATGAGAGACCAACCGATGATC 92
    || || || || || || || || || || || || || || || || || || || || ||
Db 3938167 GCGCTTGCCATTCCTCCCGCGCGGCGTTCGCGCGCTTCGCGCGCATTCGCGCGCAC 3938108
OY 93 GCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 152
    || || || || || || || || || || || || || || || || || || || || ||
Db 3938107 GCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3938048
OY 153 GAGGTTGCCAGGAGATCGCCAAATGGGTCGCCGACAGAGAGATCGCCAAATGAGAGAGCC 212
    || || || || || || || || || || || || || || || || || || || || ||
Db 3938047 GCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3937988
OY 213 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 272
    || || || || || || || || || || || || || || || || || || || || ||
Db 3937987 GGTGGCGGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3937928
OY 273 AGCTTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
    || || || || || || || || || || || || || || || || || || || || ||
Db 3937927 GCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3937868
OY 333 GGCAGAGAGATCAACATCGCGAGGCGATCGCTAAGAGAGAGAGAGAGAGAGAGAG 392
    || || || || || || || || || || || || || || || || || || || || ||
Db 3937867 GGCAGAGAGATCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3937808
OY 393 GAGGTGGCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 452
    || || || || || || || || || || || || || || || || || || || || ||
Db 3937807 CCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3937748
OY 453 GCACAGAGAGATCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 497
```

```
Db 3937747 ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 3937703
    || || || || || || || || || || || || || || || || || || || || ||
```

```
RESULT 3
US-09-197-649-7/c
Sequence 7, Application US/09197649
Patent No. 6194550
GENERAL INFORMATION:
APPLICANT: Gold, Larry
APPLICANT: Tribnow, Craig
APPLICANT: Smith, Jonathan D.
TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
FILE REFERENCE: NEX02/CI-CON
CURRENT APPLICATION NUMBER: US/09/197,649
CURRENT FILING DATE: 1998-11-23
EARLIER APPLICATION NUMBER: 07/829,461
EARLIER FILING DATE: 1992-01-31
EARLIER APPLICATION NUMBER: 07/739,055
EARLIER FILING DATE: 1991-08-01
EARLIER APPLICATION NUMBER: 07/561,968
EARLIER FILING DATE: 1990-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 390
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Sequence
OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-197-649-7
```

```
Query Match 16.6%; Score 82.8; DB 4; Length 390;
Best Local Similarity 51.3%; Pred. No. 3.7e-07;
Matches 192; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

OY 121 CCGAGAGAGGATCGAGGAGATCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
    || || || || || || || || || || || || || || || || || || || || ||
Db 388 CTGACCTTCACCAACATGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 329
OY 181 TCGCGCACAGAGGATCGCCAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
    || || || || || || || || || || || || || || || || || || || || ||
Db 328 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 269
OY 241 TCGAGGCTTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
    || || || || || || || || || || || || || || || || || || || || ||
Db 268 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 209
OY 301 GCATCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
    || || || || || || || || || || || || || || || || || || || || ||
Db 208 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 149
OY 361 CTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
    || || || || || || || || || || || || || || || || || || || || ||
Db 148 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 89
OY 421 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
    || || || || || || || || || || || || || || || || || || || || ||
Db 88 TCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 29
OY 481 CCGAGGAGATCGCC 494
    || || || || || || || || || || || || || || || || || || || || ||
Db 28 TCGTGTGTGTGTGTG 15
```

```
RESULT 4
US-09-023-173-4
Sequence 4, Application US/09023173
Patent No. 6066781
GENERAL INFORMATION:
```


Fri Oct 11 09:30:45 2002

us-09-489-101a-15_copy_1_500.rni

Page 4

QY	338	CGAGGTCACCATTCACAGAGCATTCGCTAAAGAGAGAGCCGTCACAGAGCATTCGCTAAAGAGAGT	397
Db	316	CACCGGCGGCAACGGGCGGCTTGGGCGCGCGGGGGGTGGCGGAGGCAAGCCCGCGTCCG	375
QY	398	GGCGCGCCCAAGGCGCATCGCCCAAGAGAGAGCGCCGCGCCAG	434
Db	376	TGCGGAATCCGGGCTACCATTCGACAGCGCGGCGCCAAAG	412

```

RESULT 6
US-09-130-114-2/C
Sequence 2, Application US/09130114
Patent No. 5976807
GENERAL INFORMATION:
APPLICANT: Horlick, Robert A.
APPLICANT: Damam B.
APPLICANT: Robbins, Alan K.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
FILE REFERENCE: 0867/1D9030J1
CURRENT APPLICATION NUMBER: US/09/130.114
CURRENT FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: fastseq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1931
TYPE: DNA
ORGANISM: EBNA
US-09-130-114-2

```

Query Match	15.4%	Score 76.8;	DB 2;	Length 1931;
Best Local Similarity	47.7%;	Pred. No. 4.2e-06;		
Matches 225; Conservative	0;	Mismatches 247;	Indels 0;	Gaps 0;

QY	18	GAGACGCTGCCACGGGATCGCTAAATGAGACGCCACGAGGCATCGCTAATGAGAC	77
Db	736	GACGGGAGGACGAGACGGGGAGAGACGAGAGACGGGAGAGACGAGACCGG	6777
QY	78	ACCAACCATGTGCATGGCCCAAGAGAAACGCCCCAGGGCATCGCGAGAGACGCATTCAG	1373
Db	676	GAGACGGGAGAGACGAGACCGGGAGAGACGAGACCGGGAGAGACCGGGAGAGACCGG	6177
QY	138	GGCATCGCCACGAGAGGTTGCCAGGGCATCGCCATATGGGTGCGCGACAGGGCATYC	1977
Db	616	GAGGACGAGGACGCGGAGAGAGACGAGACGGGAGAGACGCGGAGAGAGAC	5577
QY	198	GCCATATGAGGAGCGCCACCCAGGGCATCGCCACATGGGAGACCGTCCAGCGCTTCCCAAC	2577
Db	556	GAGGACGGGAGGACGCGGAGAGAGACGAGAGAGAGAGAGAGAGAGACGCGGAGAGAC	4977
QY	258	GGGAGAGCGGCTCTCAGCTTGGCCCAAGGGAGACGGCCCGCCAGGGCATCGCCACGGGAGAC	3177
Db	496	GGGAGAGCGGAGAGAGACGAGACCGGAGAGACCGGAGAGAGACGAGACCGGGAG	4377
QY	318	GCCACCAAGGGCATGSGGACAGAGTCACCATCCAGCGCATCGCTAACGAGAGACCGCTC	3777
Db	436	GAGGACGAGGAGCGGGAGAGACGGGAGAGCGGGAGAGAGAGACGAGACGGGAGAGAGAC	3777
QY	378	GAGGAGCATGCTAAGAGAGTGGCCGCCACGAGGCATGCCACGAGAGACCGCCCGGAGGA	4377
Db	376	GAGGACGGGAGAGACGGGAGAGACGGGAGAGACGGGAGAGAGAGAGACGAGACCGGAGAGAG	3177
QY	438	ATGCGCGAGAGATGTGACACAGGAGCATGCGCCAAAGAGAGACCGCGCCAGGAGCA	489
Db	316	GACGAGAGCGGGAGAGACGAGACCGGGAGAGACGAGAGAGACGAGACGAGACCA	265

RESULT 7
US-09-056-556-167
; Sequence 167, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:

```

1  APPLICANT: Reed, Steven G.
2  APPLICANT: Skeiky, Yasir A.W.
3  APPLICANT: Dillon, Davin C.
4  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
5  NUMBER OF SEQUENCES: 241
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: SEED and BERRY LLP
8  STREET: 6300 Columbia Center, 701 Fifth Avenue
9  City: Seattle
10 STATE: Washington
11 COUNTRY: USA
12 ZIP: 98104-7092
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/056,556
20 FILING DATE: 07-APR-1998
21 CLASSIFICATION:
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Maki, David J.
24 REGISTRATION NUMBER: 31,392
25 REFERENCE/DOCKET NUMBER: 210121.457
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (206) 682-4900
28 TELEFAX: (206) 682-6031
29 INFORMATION FOR SEQ ID NO.: 167:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 1439 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 IS-09-036-556-16

```

Query Match	15.2%;	Score 76.2;	DB 4;	Length 1439;
Best Local Similarity	48.9%;	Pred. No. 5.3e-06;		
Matches 235; Conservative	0;	Mismatches 243;	Indels 3;	Gaps 1;

QY	22	AGCTGACCCAGAGGAGTGCATATAGAGAGACCCAGACAGAGGATGCTGATATAGAGACACA	81
Dp	937	ACGCGACGACGACCAACCGGCAACCGCGGCTCCGGTGGGCGGGTAATAGGGGGCATTG	996
QY	82	CCCAGTGATCGGCAACGAGGAAAGC---GCCAGGGCATGCGCGAGACGACCTTCACAG	138
Dp	997	GCTTCACAGCGGCGGGGGGCGTCTGTGGGGGGCCGGGGGCGCAACGGGGTGTGTGCGGGGT	1056
QY	139	GCATGCGCAACGAGAGATTGGCCAGGCGCATTGCCATATGGGTTGCCGCAAGAGCATCG	198
Dp	1057	CCTTGGGCAACGCTGTGGGGGGGCAAGGGGGGCAACGAGGGGCAACGGGGGCGGGG	1116
QY	199	CCATATGAGAGAGCCACCCAGAGGATCGCCCACTGGAGACCCGCTTCACAGGGCTTGCCACAG	258
Dp	1117	ACGGCAACGAGGGGGGCGCCGGCGCAAGGGGGGCAACGGAGACGAGCGTGGCCGACGG	1176
QY	259	GGGAGCGCGTCTCTCAGCTTGGCCCAAGGGGAGACGGCGCCAGAGGATGCGCCACGGGGAG	318
Dp	1177	GCTCAGGGCTGTCTAACGTACATCCGCGCGCCAGAGGGGGCAACGGGGGCAATAGGGGACAG	1236
QY	319	CCACCAAGGGCATAGGCAACGAGATGCCATCACTGCATACGAGAGACGCCCTCC	378
Dp	1237	GGGGCAACGGCTCCGCGGGGGCGGCGGCAAGGGGCGTGGCGGGGAGCGCGGGAAG	1296
QY	379	AGGCGATGCTTAACGAGAGTGGCCGCCAGAGGATGCCCAAGAGAGACCGCCACAGGAA	438
Dp	1297	GGGGCAACGGGGGGGCGGCAACCGGGGGCGCCAGAGGGGCAAGGGGGCAACGGGACACGG	1356
QY	439	TGCGCGAGAGATGTGGCAACAGAGGATGGCCCAAGAGAGACCGCGCCACAGGGATGCCACA	498
Dp	1357	GTGCGCGCAAGGGGCTCAGGGGCTATCACTACGTACCTCGGCGCAAGGGGGGCAACGGGGGCA	1416
QY	499	A 499	

; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600,438US1
 ; CURRENT APPLICATION NUMBER: US/09/105,537A
 ; CURRENT FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 13613
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-105-537-3

Query Match 14.9%; Score 74.4; DB 4; Length 13613;
 Best Local Similarity 49.5%; Pred. No. 1.1e-05;
 Matches 220; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 59 GGGGATGCGTAATGAGACACACACAGTGCATGCGCCAGAGAGAGAGAGAGAGAGAGAT 118
 DB 12440 GGGGCGCCCTGCGCCGCGCGACAGCTGCGGAAAGTGCAGACAGAGAGAGAGAGAT 12381
 QY 119 GCGGAG 178
 DB 12380 GATCTTGTAGAT 12321
 QY 179 GGTGCGCGACAGAT 238
 DB 12320 GCGGAG 12261
 QY 239 GGTGCGCGAT 295
 DB 12260 GCGGAG 12201
 QY 296 CCAAGGAT 355
 DB 12200 CTTCGAT 12141
 QY 356 CATGCTTACGAT 415
 DB 12140 GCGGAT 12081
 QY 416 CAACGAT 475
 DB 12080 CAACGAT 12021
 QY 476 CGCGCGAT 499
 DB 12020 CGACGCGAT 11997

RESULT 11
 ; US-09-177-349-2/c
 ; Sequence 2, Application US/09177349
 ; Patent No. 6268201
 ; GENERAL INFORMATION:
 ; APPLICANT: Alland, David
 ; APPLICANT: Bloom, Barry R.
 ; APPLICANT: Jacobs, Jr., William R.
 ; TITLE OF INVENTION: Inhib, Inia AND Inic GENES OF MYCOBACTERIA AND METHODS
 ; FILE REFERENCE: 96700/491
 ; CURRENT APPLICATION NUMBER: US/09/177,349
 ; CURRENT FILING DATE: 1998-10-23
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 5036
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-177-349-2

Query Match 14.8%; Score 74; DB 4; Length 5036;
 Best Local Similarity 48.2%; Pred. No. 1.3e-05;
 Matches 209; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 7 AGCGATTGCTGAT 66
 DB 935 ACTGCTGCTGCTGAT 876
 QY 67 CTAATGAT 126
 DB 875 CCGGCTGAT 816
 QY 127 AGCGATTGCTGAT 186
 DB 815 CCGGCTGAT 756
 QY 187 CAGAT 246
 DB 755 CCGAT 696
 QY 247 GCTTGGCAGAT 306
 DB 695 CCGAT 636
 QY 307 CCAAGAT 366
 DB 635 CCAATGAT 576
 QY 367 AGGAT 426
 DB 575 CCGAT 516
 QY 427 CCGCGAGAGAGAT 440
 DB 515 AGAGAGAGAGAGAT 502

RESULT 12
 ; US-09-320-878-21/c
 ; Sequence 21, Application US/09320878A
 ; Patent No. 6117659
 ; GENERAL INFORMATION:
 ; APPLICANT: ASHLEY, Gary
 ; APPLICANT: BETLACH, Melanie C.
 ; APPLICANT: MCDANIEL, Robert
 ; APPLICANT: TANG, Li
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
 ; FILE REFERENCE: 30662002120
 ; CURRENT APPLICATION NUMBER: US/09/320,878A
 ; EARLIER FILING DATE: 1999-05-27
 ; EARLIER FILING DATE: CIP OF 09/141,908
 ; EARLIER FILING DATE: 1998-08-28
 ; EARLIER FILING DATE: CIP OF 09/073,538
 ; EARLIER FILING DATE: 1998-05-06
 ; EARLIER FILING DATE: CIP OF 08/846,247
 ; EARLIER FILING DATE: 1997-04-30
 ; EARLIER FILING DATE: 1999-02-08
 ; EARLIER FILING DATE: 1999-02-08
 ; EARLIER FILING DATE: 1998-09-22
 ; EARLIER FILING DATE: 1998-09-22
 ; EARLIER FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 21
 ; LENGTH: 5970
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-320-878-21

Query Match 14.8%; Score 74; DB 3; Length 5970;
 Best Local Similarity 49.3%; Pred. No. 1.3e-05;

Matches 219; Conservative 1; Mismatches 221; Indels 3; Gaps 1;

QY	59	GGGCAATCGGTAAATGAGAAACACACCCAGTGCATTCGCAACGAGGAAAGCCGCCAGGGCAT	118
Pb	552	GGGCGCCCTCTGGCCGCCGCAACGACTGTGGAAAGTTCGCGCAAGCAAGGACCTCTGGGCT	493
QY	119	CGCCGAGAGACGCCATTCAGGGGCATTCGCCAACGAGAGAGTTGGCCAGGGGCATTCGCATGG	178
Pb	492	GTACTTCCTGAGCGCCGCGCACGCCCTCGGTGCGGGGTGTGAGAGGGCCGGGCCCGCGGCACCT	433
QY	179	GATGCGCCGACACAGGGGCATTCGCAATATGAGAGCGCCACCCAGGGGCATTCGCAATGGAGCG	238
Pb	432	CGCGCGACGCGGAGGTCTTTCACATTCCTCCAGCGCAACAGGCCGCTTCAACGCCCTTCGAGGGCGG	373
QY	239	CGTTCACAGGCTTTCGCGCAACGGGGGA---CGCCGTCCTCACTTCGCGCAACGGGGAGCGCGC	295
Pb	372	CGCCGTCGTACCGAAGAGAGGCCGACCTTCGCGCCCGCGGATTCGCGCCCTTCACAACTTTCGG	313
QY	296	CCAGGGGCATTCGCGCAACGGGGAGAGCGCCACCAAGGGGCATTCGGGCAACGAGAGTCAACATTCACGG	355
Pb	312	CTTTCGACTCTGGCCCGGCGCGGCGAGCCCCCGCGCGGGAGCAACGCCCAAGATGAGAGGCGCG	253
QY	356	CATTCGCTAAACGAGAGAGCGCCGTCACAGGGGCATTCGCTAAACGAGGTGGCGCGCCAGGGGCATTCG	415
Pb	252	CGCGCGCATAGGGCCCTTCACCTTCCTTCGAGAGGGTTCCCGAGAGTCAATCGACCGGAACCGCGC	193
QY	416	CAACGAGAGAGCGCGCGCCAGAGGAATTCGCGAGAGAGTTCGACACAGGGGCATTCGCAACGAGGA	475
Pb	192	CACACAGAGCGCTCAACCGCGAGCACTTCGCGAGACCTCCCGCGCTCTCTGTGCGCGACCA	133
QY	476	CGCGCGCCAGAGGCATTCGCGCAACA	499
Pb	132	CGACCGCGACGGGCTTCAACACCA	109

RESULT 13 942C-1
US-08-387-942C-1
: Sequence 1, Application US/08387942C
: Patent No. 5939289
: GENERAL INFORMATION:
: APPLICANT: ERTESVAG, HELGA
: APPLICANT: VALLA, SVEIN
: APPLICANT: SKJAK-BRAEK, GUDMUND
: APPLICANT: LARSEN, BJORN
: TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
: TITLE OF INVENTION: ENCODING MANNIRONAN C-5-EPIMERASE
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESS: BIRCH, STEWART, KOLASCH & BIRCH, LLP
: STREET: P.O. BOX 747
: CITY: FALLS CHURCH
: STATE: VA
: COUNTRY: USA
: ZIP: 22042
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/387,942C
: FILING DATE: 09-MAY-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURPHY JR, GERALD M.
: REGISTRATION NUMBER: 28, 977
: REFERENCE/DOCKET NUMBER: 1809-106P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8030
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12588 base pairs

```

?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      ORIGINAL SOURCE:
?      ORGANISM: Azotobacter vinelandii
?      STRAIN: E
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 290..1951
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 2227..6438
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 6702..9695
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 9973..12588
US-08-387-942C-1

```

[illegible]

: CURRENT APPLICATION NUMBER: US/09/105,537A
 : CURRENT FILING DATE: 1998-06-26
 : NUMBER OF SEQ ID NOS: 43
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 1
 : LENGTH: 15872
 : TYPE: DNA
 : ORGANISM: Streptomyces venezuelae
 : US-09-105-537-1

Query Match 14.6%; Score 73.2; DB 4; Length 15872;
 Best Local Similarity 47.2%; Pred. No. 1.8e-05;
 Matches 222; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 8 CGGCATTGCTGAGGACGCTGCCAGGCGCATGCTAATGAGCGCCGACAGGCGCATCCG 67
 Db 4792 CGCCATGGGCGAGAGACTCGCGCGCGCGACCCCGTGTTCGCGCGCGCCCTCGACCGT 4851
 QY 68 TAATGAGGACACCAACCACTGATGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 127
 Db 4852 GTACGGGGGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 4911
 QY 128 CGCCATCGAGGCGATGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 187
 Db 4912 GAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 4971
 QY 188 ACAGGCGATCGCCATGAG 247
 Db 4972 CGCCATCGAGGCGATGAG 5031
 QY 248 CTTGCGCAACGGGAG 307
 Db 5032 GATCGCGCGCGCGAG 5091
 QY 308 CAACGGGAG 367
 Db 5092 CGCCCGGCGCGCGAG 5151
 QY 368 GAGAGCGCTGACAG 427
 Db 5152 GAGAGCGAG 5211
 QY 428 CGCCCGAG 477
 Db 5212 CGTCAACGAG 5261

RESULT 15

: US-09-025-691-2
 : Sequence 2, Application US/09025691
 : Patent No. 6069299

GENERAL INFORMATION:

: APPLICANT: Broadway, Roxanne M.
 : APPLICANT: Harman, Gary E.
 : TITLE OF INVENTION: FUNGUS AND INSECT CONTROL WITH
 : TITLE OF INVENTION: CHITINOLYTIC ENZYMES
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 : STREET: Clinton Square, P.O. Box 1051
 : CITY: Rochester
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 14603
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/025,691
 : FILING DATE:
 : CLASSIFICATION:

: ATTORNEY/AGENT INFORMATION:
 : NAME: Goldman, Michael L.
 : REGISTRATION NUMBER: 30,727
 : REFERENCE/DOCKET NUMBER: 19603/20120
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (716) 263-1304
 : TELEFAX: (716) 263-1600
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1294 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : US-09-025-691-2

Query Match 14.5%; Score 72.6; DB 3; Length 1294;
 Best Local Similarity 52.0%; Pred. No. 2.3e-05;
 Matches 193; Conservative 0; Mismatches 169; Indels 9; Gaps 1;

QY 128 CGCCATCGAGGCGATGCCAG 187
 Db 317 CACCGGCGCGGAG 376
 QY 188 ACAGGCGATCGCCATGAG 247
 Db 377 ACCGGCGAG 436
 QY 248 CTTGCGCAACGGGAG 307
 Db 437 CTTGCGCGAG 496
 QY 308 CAACGGGAG 367
 Db 497 CAACGGGAG 556
 QY 368 GAGAGCGCTGACAG 427
 Db 557 G-----CGATTGCGGCGAG 607
 QY 428 CGCCCGAG 487
 Db 608 CATCTGCGTGGCGGAG 667
 QY 488 CATCGGCAACA 498
 Db 668 CTTGCGCGACA 678

Search completed: October 10, 2002, 19:46:20
 Job time: 1097.73 secs

Fri Oct 11 09:30:46 2002

us-09-489-101a-15_copy_1_500.rst

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds
(without alignments)
9699,805 Million cell updates/sec

Title: US-09-489-101a-15_COPY_1_500

Perfect score: 500

Sequence: 1 gagctacgcatgctgag.....ccacggcagcagcacaag 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 2747244

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estha:*
2: em_esthu:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_ncc:*
9: g0_estc1:*
10: g0_estc2:*
11: g0_hcc:*
12: g0_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	153.4	30.7	697	9	AL110383
2	124.6	24.9	179	9	AL041090
3	118.6	23.7	758	10	BI829297
4	113	22.6	673	10	BI561135
5	98.6	19.7	582	10	BG724330
6	94.2	18.8	241	9	AI904151
7	85.4	17.1	780	10	BG847089
8	80.8	16.2	822	12	CNS0458K
9	79.6	15.9	640	10	BE427000
10	79.2	15.8	884	10	BG417341
11	78.4	15.7	990	9	BE216934
12	76.6	15.3	602	9	AI176645
13	74.8	14.8	681	12	CNS0280D
14	73.8	14.7	915	12	CNS03P82
15	73.4	14.7	401	9	AV438522
16	70.2	14.0	511	10	BG739444
17			511	10	BG773077

18	69.8	14.0	454	9	BE125156
19	69.6	13.9	673	10	BG307710
20	69.2	13.8	538	10	BM178969
21	69.2	13.8	679	9	AL506812
22	69.2	13.8	814	10	BE260740
23	68.6	13.7	554	10	BG367845
24	68	13.6	530	10	BF483460
25	68	13.6	689	9	AU076299
26	67.8	13.6	619	10	BG300721
27	67.6	13.5	564	10	BE192599
28	67.6	13.5	652	10	BG314481
29	67.4	13.5	321	10	BG874723
30	67.4	13.5	938	9	BE213813
31	67.2	13.4	925	12	CNS0091P
32	67.2	13.4	704	9	AU088578
33	67	13.4	461	10	BE517259
34	66.8	13.4	658	9	AU088571
35	66.8	13.3	456	9	AU222996
36	66.6	13.3	529	10	BF474993
37	66.6	13.3	529	10	BE470936
38	66.6	13.3	676	9	AV916390
39	66.6	13.3	676	9	AV916390
40	66.4	13.3	633	9	AV941861
41	66.2	13.2	596	10	BE360540
42	66.2	13.2	770	10	BF265823
43	66	13.2	459	10	BE605180
44	65.6	13.1	401	9	AV438522
45	65.6	13.1	537	10	BG904964

ALIGNMENTS

RESULT 1
AL110383 697 bp mRNA linear EST 29-FEB-2000
DEFINITION DKEP434E0731.1 434 (synonym: hncs3) Homo sapiens cDNA clone
LOCUS DKEP434E0731.1 5', mRNA sequence.
ACCESSION AL110383
VERSION AL110383.1 GI:5865991
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 697)
AUTHORS Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE EST (Blum, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Blum H

MFPS
Am Klopferplatz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No 1 sequence available.
This clone (DKEP434E0731) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1..697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKEP434E0731"
/clone_id="434 (synonym: hncs3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pSport1; Site_1: Not; Site_2: SalI"

BASE COUNT
217 a 181 c 180 g 119 t


```

Db 220 CTGACGACTCTGACGACGCTCCGGCTGTGAAGAGGGCCCTCCGCAAGGCCCTGCCA 279
OY 76 ACACACACCCAGGCAATCCGCAACGAGAACCCCGGAGGCAATCCGCAAGGCCCTGCC 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 AGAAGAGAGATCTCTCCACGACTCGACACACACACACACACACACACACACACAC 339
OY 136 AGGCGATCCGCAACGAGAGGTTCGCCAGGCAATCCGCAATGGGTTCGCCACACAGGCA 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 ACAGACCCCGCAAGAGGGGCTCCCGCCCGCAAGGCCCTCCGCAAGAGAGGCCGAC 399
OY 196 TCCGCAATGAGAGGACCAAGGCAATCCGCAATGGAGCGCGTCCAGCGCTTGCCCA 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 GCACACACATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGAT 459
OY 256 ACAGGAGACCGCTCTCTGACTTCCGCAACGAGGAGCCCGCCGAGGCAATCCGCAAGCGG 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 AGGCAAGCCCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 519
OY 316 ACAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAAT 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 AGTGTCCGACGACGACGACGATGATGATGATGATGATGATGATGATGATGATGAT 576
OY 376 TCCGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAATGAGGCAAT 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 ---AGGCGCCCAAGAGAGGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 633
OY 436 GAATGCGCCGAGGATGCGCAGAGGCAATGCGCAGAGAGGCGCCCGCAGGCGCAATG 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 ACAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 690

RESULT 8
CNS04SGK/c 822 bp DNA linear GSS 24-MAY-2000
LOCUS Definition 007M14 of library H from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL305165.1 GI:8196096
VERSION GSS: genome survey sequence.
KEYWORDS Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetradon.
1 (bases 1 to 822)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quelier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Unpublished
2 (bases 1 to 822)
Roest-Crolius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizes,C., Mincker,P., Brotlier,P., Quelier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 822)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
Location/Qualifiers
1..822
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="007M14"

BASE COUNT 147 a 262 c 203 g 209 t 1 others
ORIGIN
Query Match 16.2% Score 80.8; DB 12; Length 822;
Best Local Similarity 50.9%; Pred. No. 0.021; Indels 3; Gaps 1;
Matches 217; Conservative 0; Mismatches 206;
Query 20 GGACCTTCCCGAGGCAATCCGTAATGAGAGCGGACGAGGCAATGCTAATGAGACAC 79
Db 556 GGAATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 497
OY 80 CACCCAGTGCATCCCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 AGGCTAATGCTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 437
OY 137 GGGCATGCGCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 AGTATGCGCTAACCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 377
OY 197 CGCCTAATGAGAGCGGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CGCTAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 317
OY 257 CGGAGAGCGGCTCTGAGCTTGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 CCGAGAGTCAAGCTAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 257
OY 317 CGGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 GTCAGGCTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 197
OY 377 CAGAGGATGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 CTAAGTATGCTAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 137
OY 437 AATGCG 442
    ||| |||
Db 136 TCAAGC 131

RESULT 9
BE427000 640 bp mRNA linear EST 24-JUL-2000
LOCUS Definition BE427000 clone PSR6561, mRNA sequence.
ACCESSION BE427000
VERSION BE427000.1 GI:9424933
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
1 (bases 1 to 640)
Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemont,J.M., Jia,J., Joudrier,P.,
Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogilhare,Y.,
Perchion,N., Qualset,C., Schuch,W., Selvaraj,G., Sharifou,M.,
Sorrells,M., Warburton,M. and Wenzel,G.
International Triticale EST Cooperative (ITREC): Production of
Expressed Sequence Tags for Species of the Triticace
Unpublished (2000)
Contact: Bailey P
Cereals Group, John Innes Centre
Norfolk, Norwich NR4 7UH UNITED KINGDOM
Tel: 44 1603 452571 ext. 2587
Fax: 44 1603 502241
Email: paul.bailey@bbsrc.ac.uk
International Triticale EST Cooperative (ITREC)

```


OY 255 AACGGGAGCCCTCTCTACGCTTCCGCAACGGGAGCCGCCGCGCATCGCCACAGG 314
 DB 348 AACCTTGAGGAGCCCTCTCTACGCTTCCGCAACGGGAGCCGCCGCGCATCGCCACAGG 407
 OY 315 GAGCGCCACGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 374
 DB 408 GAGCGCCGCGCGCATGCTCGAGAGCCGCTCGCGCATCGCTCGCTCGCTCGCTCGAC 467
 OY 375 GTCAGGAGGATCGCTAACGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 434
 DB 468 GTCGAGGATCGCTAACGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 527
 OY 435 GGAATGCGGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 462
 DB 528 GCAACCTGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 555

RESULT 11
 BE216934 990 bp mRNA linear EST 03-JUL-2000
 LOCUS BE216934
 DEFINITION EST0477 Triticum aestivum Lambda Zap Triticum aestivum cDNA clone
 JAL_5B_H05_T3 5', mRNA sequence.
 BE216934
 BE216934.1 GI:8904620
 EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 : Triticaceae; Triticum.
 1 (bases 1 to 990).

REFERENCE
 AUTHORS Anderson J.M., Williams C.E. and Goodwin S.B.
 TITLE Analysis of an EST database reveals a probable Cf2 resistance gene
 JOURNAL Unpublished (2000)
 COMMENT Crop Production & Pest Control Research Unit
 USDA-ARS
 1150 Lilly Hall, West Lafayette, IN 47907, USA
 Tel: 765-494-5565
 Fax: 765-496-2926
 Email: janderson@purdue.edu
 Seq primer: T3
 High quality sequence stop: 990.

FEATURES
 source
 1. 990
 /organism="Triticum aestivum"
 /strain="p29"
 /db_xref="taxon:4565"
 /clone="JAL_5B_H05_T3"
 /clone_lib="Triticum aestivum Lambda Zap"
 /tissue_type="leaf"
 /dev_stage="9 day old seedlings"
 BASE COUNT 182 a 360 c 290 g 158 t
 ORIGIN

Query Match 15.7%; Score 78.4; DB 9; Length 990;
 Best Local Similarity 50.0%; Pred. No. 0.043;
 Matches 230; Conservative 0; Mismatches 221; Indels 9; Gaps 1;

OY 3 GTCAGGAGGATCGCTAACGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 62
 DB 141 GCTATATCG 200
 OY 63 ATGCGTATGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 122
 DB 201 ATGAGGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCG 260
 OY 123 GAGGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
 DB 261 GCGGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 311
 OY 183 GCGGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 242

DB 312 CTCAGCG 371
 OY 243 CACGCTTGCG 302
 DB 372 TGGCGTGGCG 431
 OY 303 ATGCGCAAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 362
 DB 432 CGCAGGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCG 491
 OY 363 AACGAGGAGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCG 422
 DB 492 GTCATGGGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 551
 OY 423 GAGCG 462
 DB 552 AAGAGCATGCG 591

RESULT 12
 AUI76645 602 bp mRNA linear EST 21-MAR-2001
 LOCUS AUI76645
 DEFINITION AUI76645 Sugano-Kawakami 5' end enriched cDNA library (OLA) from
 HNI Oryzias latipes cDNA clone OLA14.06c similar to P1715282641
 hydroxyproline-rich glycoprotein - maize, mRNA sequence.
 AUI76645
 AUI76645.1 GI:13424771
 EST.
 SOURCE Japanese medaka.
 ORGANISM Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthopterygii; Acanthopterygii; Perciformes; Atherinomorpha;
 Belontiiformes; Atherinomorpha; Oryziatidae; Oryziatidae; Oryzias.
 1 (bases 1 to 602)

REFERENCE
 AUTHORS Naruse, K., Mitani, H. and Tanaka, M.
 TITLE Medaka EST project in University of Tokyo (2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Kiyoshi Naruse
 Department of Biological Sciences
 Graduate School of Science, University of Tokyo
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
 Tel: 81-3-5841-4443
 Fax: 81-3-5841-4410
 Email: naruse@biol.s.u-tokyo.ac.jp
 This clone was isolated from Sugano-Kawakami 5' end enriched cDNA
 library (OLA).

FEATURES
 source
 1. 602
 /organism="Oryzias latipes"
 /strain="HNI"
 /db_xref="taxon:8090"
 /clone="OLA14.06c"
 /clone_lib="Sugano-Kawakami 5' end enriched cDNA library
 (OLA) from HNI"
 /tissue_type="whole body"
 /dev_stage="adult"
 BASE COUNT 132 a 246 c 175 g 49 t
 ORIGIN

Query Match 15.3%; Score 76.6; DB 9; Length 602;
 Best Local Similarity 49.5%; Pred. No. 0.078;
 Matches 227; Conservative 0; Mismatches 229; Indels 3; Gaps 1;

OY 29 CAGGAGGATCGCTAACGAGGATGGGCAAGAGATCAGCATCCAGCGCATCGCTAACGAGAGCC 88
 DB 119 CAGGAGGATCGCTAACGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCG 178
 OY 89 CAGGAGGATCGCTAACGAGGATGGGCAAGAGATCAGCATCCAGCGCGCGCGCGCGCGCGCG 145
 DB 179 TGTGGCG 238

QY 146 CAAGAGAGAGTTGCCAGGCGCATGCGCAATGGGGTCCGCCACAGGGCATTCGCCATCA 205
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 239 CGAGCGCCGACGACACCACTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 298
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 206 GGAGCGCCAGGCGCATGCG 265
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 299 CGCCAGACACCACTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 358
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 266 CGTCCAGGCG 325
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 359 CACCAACACTGTGGCG 418
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 326 GGGCATGGGCG 385
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 419 GACTGTGGCG 478
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 386 CGCTAACGAGGTGGCG 445
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 479 GCGCGAGCG 538
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 446 GGATGTGCGACAGGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 484
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 539 CGCCGACGACACCACTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 577
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 RESULT 13
 CDS02E0D/LOCUS 681 bp DNA linear GSS 13-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 262H14 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL193990.1 GI:7632096
 VERSION AL193990
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 681)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 681)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 681)
 Genome.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.
 Location/Qualifiers
 1..681
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="262H14"
 /clone_11b="G"
 /note="Genoscope sequence ID : C0AG262DD07LPI-end : 'T7"
 BASE COUNT 74 a 105 c 170 g 282 t 50 others
 ORIGIN
 Query Match 15.0%; Score 74.8; DB 12; Length 681;

Best Local Similarity 45.4%; Pred. No. 0.13;
 Matches 173; Conservative 21; Mismatches 187; Indels 0; Gaps 0;
 QY 14 TGGTAGAGAGCGTGTCCCGAGGGCATGCTTAATGAGAGCGCGCGCGCGCGCGCGCGCG 73
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 412 TTTTGTCDACRACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 353
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 74 GGACACCAACCCAGTGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 133
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 352 CGACACACACGACGACRACGACGACGACGACGACGACGACGACGACGACGACGACGAC 293
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 134 CCAGGCGCATGCG 193
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 292 CACG 233
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 194 CATGCGCATGAGAGCG 253
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 232 CACCAAC 173
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 254 CACG 313
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 172 CAAAAACRACAAACRACRACRACRACRACRACRACRACRACRACRACRACRACRAC 113
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 314 GGACG 373
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 112 CACCAAC 53
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 QY 374 CGTCCAGGCG 394
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 Db 52 CAAAC 32
 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
 RESULT 14
 CDS03P82
 LOCUS 915 bp DNA linear GSS 17-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-01 end of clone
 044F19 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL254315.1 GI:7975327
 VERSION AL254315
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 915)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 915)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 915)
 Genome.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.
 Location/Qualifiers
 1..915
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 FEATURES
 source

/clone="044f19"
/clone_1lb="g"
/note="Genoscope sequence ID : C09C044C10SP1-end :
PUC-0.1"

BASE COUNT 317 a 245 c 208 g 140 t 5 others

Query Match 14.8%: Score 73.8; DB 12; Length 915;
Best Local Similarity 54.9%; Pred. No. 0.18;
Matches 168; Conservative 0; Mismatches 133; Indels 5; Gaps 1;

QY 194 CATGCCATATAGAGAGCCACCCAGGCGATCGCCACTGGAGAGCCGCTCAGGCTTCC 253
DB 394 CAACGACACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 453
QY 254 CAACGGGAGACGCGCTCTCAGCTTCCGCAACGGGAGCGCCGCGGATCGCCACGCG 313
DB 454 CGACGACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 513
QY 314 GGAGCGCCACAGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 373
DB 514 TGACACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 568
QY 374 CGTCCAGGCGATCGCTTACGAGGTGCGCCGCGGCGATCGCCACGACGACGCGCCCA 433
DB 569 CGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 628
QY 434 GGGATCGCGGAGTGTGCGACAGGCGATCGCCACGACGACGACGACGACGACGACGACG 493
DB 629 CAACGGCAACAGCAGCGGCGACGACGACGACGACGACGACGACGACGACGACGACG 688
QY 494 CAACAA 499
DB 689 CAACAA 694

RESULT 15
AVA38522/c 401 bp mRNA linear EST 23-AUG-2000
LOCUS AVA38522 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION PS037403_r 5', mRNA sequence.
ACCESSION AVA38522
VERSION AVA38522.1 GI:8593747
KEYWORDS EST.
SOURCE Porphyra yezoensis.
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.

REFERENCE 1 (bases 1 to 401)
AUTHORS Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
Tabata, S.
TITLE Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 401
/organism="Porphyra yezoensis"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PS037403_r"
/clone_1lb="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 37 a 187 c 124 g 53 t

Query Match 14.7%: Score 73.4; DB 9; Length 401;

Best Local Similarity 51.8%; Pred. No. 0.22;
Matches 192; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

QY 122 CGAGAGCGCCATTCAGGCGATCGCCACAGAGAGGTGCCAGGCGATTCGCCAATGGGGT 181
DB 372 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 313
QY 182 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241
DB 312 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 253
QY 242 CGACGGCTTCGCCACGCG--GGAGCGGCTCTCAGCTTCCGCAACGGGAGCGGCGGCG 298
DB 252 GGGCAGGTACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 193
QY 299 GGGCATCGCCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 358
DB 192 GGGCAGGTACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 133
QY 359 CGCTAACGAGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 418
DB 132 CGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 73
QY 419 CGAGAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 478
DB 72 CGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13
QY 479 CGGCCAGGGCA 489
DB 12 CGCGGTGGGAA 2

Search completed: October 10, 2002, 20:28:27
Job time : 701.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 84.9333 Seconds

(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-15_COPY_1_500

Sequence: 1 gagctcagcgatgctgag.....ccacggcgatcgccaacaag 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N_Geneseq_0132802.*

1: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2000.DAT.*
21: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2001A.DAT.*
22: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2001B.DAT.*
23: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2001C.DAT.*
24: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	748	22	AA11123 Human small cell 1
2	500	100.0	1201	22	AA11122 Human small cell 1
3	307.8	61.6	1593	23	AA572677 DNA encoding novel
4	302.4	60.5	3093	23	AA572678 DNA encoding novel
5	269.6	53.9	2922	23	AA572679 DNA encoding novel
6	257.6	51.5	518	22	AA66861 Human foetal liver
7	257.6	51.5	518	22	AA633930 Probe #12396 for g
8	257.6	51.5	518	22	AA633930 Human brain expres
9	257.6	51.5	518	22	AA147072 Probe #15758 used

10	250	50.0	3345	23	AA572676 DNA encoding novel
11	250	50.0	3522	23	AA572321 DNA encoding novel
12	235	47.0	495	22	AB542299 Human foetal liver
13	235	47.0	495	22	AB542056 Probe #2522 for ge
14	235	47.0	495	22	AB542056 Human brain expres
15	235	47.0	495	22	AA133949 Probe #2635 used t
16	181.4	36.3	1530	21	AA166669 Human secreted pro
17	178.2	35.6	537	22	AAK22452 Human brain expres
18	178.2	35.6	537	22	AAK22452 Human bone marrow
19	178.2	35.6	537	22	AA124443 Probe #14376 for g
20	178.2	35.6	537	22	AA124443 Probe #23134 used
21	178.2	35.6	576	22	AA561447 Human breast cell
22	178.2	35.6	593	22	AAK08801 Probe #12924 for g
23	178.2	35.6	593	22	AAK35695 Human brain expres
24	178.2	35.6	593	22	AAK35695 Human bone marrow
25	178.2	35.6	593	22	AAK35695 Human bone marrow
26	171	34.2	586	22	AB560111 Human foetal liver
27	171	34.2	586	22	AAK08387 Human brain expres
28	171	34.2	586	22	AAK34267 Human bone marrow
29	169.2	33.8	468	22	AA139989 Probe #8675 used t
30	169.2	33.8	468	22	AA15206 Probe #5050 for ge
31	153.6	30.7	471	22	AB572651 Human foetal liver
32	153.6	30.7	471	22	AAK21080 Human foetal expres
33	153.6	30.7	471	22	AAK47234 Human bone marrow
34	153.6	30.7	471	22	AAI53069 Probe #21755 used
35	145.4	29.1	3127	23	AA576376 DNA encoding novel
36	145.4	28.5	471	22	AB56456 Probe #14922 for g
37	142.4	28.5	471	22	AB56456 Probe #14318 for g
38	140.8	28.2	459	22	AB545009 Human breast cell
39	140.8	28.2	459	22	AA113771 Probe #3704 for ge
40	140.8	28.2	459	22	AA113771 Human foetal liver
41	140.8	28.2	526	22	AB560045 Human brain expres
42	140.8	28.2	526	22	AAK08319 Human bone marrow
43	140.8	28.2	526	22	AAK34197 Probe #8606 used t
44	136	27.2	526	22	AA61896 Human foetal liver

ALIGNMENTS

RESULT 1	AA11123 standard; DNA: 748 BP.
ID	AA11123
AC	AA11123
DT	24-SEP-2001 (first entry)
XX	Human small cell lung cancer associated gene, DKFZp434C196 #2.
XX	Human; small cell lung cancer; therapy: hCNAp; nucleic acid: NA;
XX	melanoma; cancer; colon; breast; head; neck; transitional cancer;
XX	leiomyosarcoma; synovial sarcoma; cytostatic; ds.
XX	Homo sapiens.
XX	WO200153349-A2.
XX	26-JUL-2001.
XX	19-JAN-2001; 2001MC-DS02015.
XX	PF 21-JAN-2000; 2000US-0489101.
XX	PR (LUDW-) LUDWIG INST CANCER RES.
XX	PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX	PA (CORR) CORNELL RES FOUND INC.
XX	Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX	WPI, 2001-457597/49.

PT Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e-9. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 XX
 PS Claim 57; Page 116; 152pp; English.

XX The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterized by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterized by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterized by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is human DKFZp434C196 DNA. This small cell lung
 CC cancer associated gene is designated as NY-SCIC-13.
 CC
 XX

SQ Sequence 748 BP; 166 A; 261 C; 246 G; 72 T; 3 other;

Query Match 100.0%; Score 500; DB 22; Length 748;
 Best Local Similarity 100.0%; Pred. No. 1.4e-67;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCTCTACGGCATTGCTGAGAGAGGCTGCGGATCGCTAATGAGAGCGGCGACGAG 60
 DB 1 GAGCTCTACGGCATTGCTGAGAGAGGCTGCGGATCGCTAATGAGAGCGGCGACGAG 60
 QY 61 GCATCGCTATGAG 120
 DB 61 GCATCGCTATGAG 120
 QY 121 CCGAG 180
 DB 121 CCGAG 180
 QY 181 TCGGCGGAG 240
 DB 181 TCGGCGGAG 240
 QY 241 TCGAGGCGCTTGGCCCAAGGGGAGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAG 300
 DB 241 TCGAGGCGCTTGGCCCAAGGGGAGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAG 300
 QY 301 GCATCGGCAAGGGGAGCGGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAGCG 360
 DB 301 GCATCGGCAAGGGGAGCGGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAGCG 360
 QY 361 CTAAAGGAG 420
 DB 361 CTAAAGGAG 420
 QY 421 AGGAG 480
 DB 421 AGGAG 480
 QY 481 CCCAGGCGCATCGCCCAAG 500
 DB 481 CCCAGGCGCATCGCCCAAG 500

RESULT 2
 AAD11122
 ID AAD11122 standard; DNA: 1201 BP.
 XX
 AC AAD11122;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE Human small cell lung cancer associated gene, DKFZp434C196 #1.
 XX

KW Human; small cell lung cancer; therapy; hCAP; nucleic acid; NA;
 KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
 KW leiomyosarcoma; synovial sarcoma; cytostatic; ds.
 OS Homo sapiens.
 XX
 XX
 PN WO200153349-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 19-JAN-2001; 2001WO-US02015.
 XX
 PR 21-JAN-2000; 2000US-0489101.
 XX
 PR (LUDWIG) INST CANCER RES.
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
 XX
 DR WPI; 2001-457597/49.

PT Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e-9. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 XX
 PS Claim 57; Page 114-115; 152pp; English.

CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterized by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterized by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterized by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is human DKFZp434C196 DNA. This small cell lung
 CC cancer associated gene is designated as NY-SCIC-13.
 CC
 XX

SQ Sequence 1201 BP; 255 A; 377 C; 349 G; 178 T; 42 other;

Query Match 100.0%; Score 500; DB 22; Length 1201;
 Best Local Similarity 100.0%; Pred. No. 1.3e-67;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCTCTACGGCATTGCTGAGAGAGGCTGCGGATCGCTAATGAGAGCGGCGACGAG 60
 DB 1 GAGCTCTACGGCATTGCTGAGAGAGGCTGCGGATCGCTAATGAGAGCGGCGACGAG 60
 QY 61 GCATCGCTATGAG 120
 DB 61 GCATCGCTATGAG 120
 QY 121 CCGAG 180
 DB 121 CCGAG 180
 QY 181 TCGGCGGAG 240
 DB 181 TCGGCGGAG 240
 QY 241 TCGAGGCGCTTGGCCCAAGGGGAGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAG 300
 DB 241 TCGAGGCGCTTGGCCCAAGGGGAGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAG 300
 QY 301 GCATCGGCAAGGGGAGCGGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAGCG 360
 DB 301 GCATCGGCAAGGGGAGCGGCGGCTCTCAGCTTCGCGCAAGGGGAGCGGCGAGCG 360
 QY 361 CTAAAGGAG 420

XX WP1: 2001-48899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 PS Claim 4: SEQ ID No 12396; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 518 BP; 140 A; 170 G; 139 G; 69 T; 0 other;
 Query Match 51.5%; Score 257.6; DB 22; Length 518;
 Best Local Similarity 75.2%; Pred. No. 5.1e-31;
 Matches 376; Conservative 0; Mismatches 114; Indels 10; Gaps 4;
 XX
 XX 5 CTACGGCATTTGCTGAGAGCGTGTCCAGGCGATGCTAATAGAGAGAGCGGCGGCGAT 64
 XX 11
 XX 23 CAACATCACTAAGAGAGATGCTTACAGATATCACTAATAGAGAGAGCGGCGGCGAT 82
 XX 11
 XX 65 CGCTAATGAG 121
 XX 11
 XX 83 CGCTAATGAG 142
 XX 11
 XX 122 CGAG 181
 XX 11
 XX 143 CGAG 202
 XX 11
 XX 182 CGCGAG 241
 XX 11
 XX 203 CGCGAG 262
 XX 11
 XX 242 CACAGGCTTCCGCAAG 301
 XX 11
 XX 263 CCAATGATCGGCAAG 322
 XX 11
 XX 302 CATTGCAAG 361
 XX 11
 XX 323 CATGCTAATGAG 382
 XX 11
 XX 362 TAACGAG 421
 XX 11
 XX 383 TA---ATGACACCGCTACAGAGCA -CGCTAAGAGAGAGAGAGAGAGAGAGAGAG 438
 XX 11
 XX 422 GAG 478
 XX 11
 XX 439 GAG 498
 XX 11
 XX 479 CGCCGAG 498
 XX 11
 XX 499 TCTACAG 518
 XX 11
 XX
 XX RESULT 8
 XX AAK15297
 XX ID AAK15297 standard; DNA; 518 BP.
 XX AC AAK15297;
 XX DT 05-NOV-2001 (first entry)
 XX Human brain expressed single exon probe SEQ ID NO: 15288.

XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO200157275-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00667.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632465.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WP1: 2001-483446/52.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 XX brains -
 XX Example 4: SEQ ID NO: 15288; 650pp + Sequence listing; English.
 XX
 XX The present invention provides a number of single exon nucleic acid
 XX probes which are derived from genomic sequences expressed in the human
 XX brain. They can be used to measure gene expression in brain cell systems,
 XX which may enable the diagnosis and improved treatment of nervous system
 XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 XX epilepsy and cancers. The present sequence is one of the probes of the
 XX invention.
 XX
 XX Sequence 518 BP; 140 A; 170 G; 139 G; 69 T; 0 other;
 Query Match 51.5%; Score 257.6; DB 22; Length 518;
 Best Local Similarity 75.2%; Pred. No. 5.1e-31;
 Matches 376; Conservative 0; Mismatches 114; Indels 10; Gaps 4;
 XX
 XX 5 CTACGGCATTTGCTGAGAGCGTGTCCAGGCGATGCTAATAGAGAGAGCGGCGGCGAT 64
 XX 11
 XX 23 CAACATCACTAAGAGAGATGCTTACAGATATCACTAATAGAGAGAGCGGCGGCGAT 82
 XX 11
 XX 65 CGCTAATGAG 121
 XX 11
 XX 83 CGCTAATGAG 142
 XX 11
 XX 122 CGAG 181
 XX 11
 XX 143 CGAG 202
 XX 11
 XX 182 CGCGAG 241
 XX 11
 XX 203 CGCGAG 262
 XX 11
 XX 242 CACAGGCTTCCGCAAG 301
 XX 11
 XX 263 CCAATGATCGGCAAG 322
 XX 11
 XX 302 CATTGCAAG 361
 XX 11
 XX 323 CATGCTAATGAG 382
 XX 11
 XX 362 TAACGAG 421
 XX 11
 XX 383 TA---ATGACACCGCTACAGAGCA -CGCTAAGAGAGAGAGAGAGAGAGAGAGAG 438
 XX 11

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 3345 BP; 954 A; 940 C; 913 G; 538 T; 0 other;

Query Match 50.0%; Score 250; DB 23; Length 3345;
 Best Local Similarity 74.1%; Pred. No. 5, 6e-30;

Matches 372; Conservative 0; Mismatches 120; Indels 10; Gaps 4;

OY 5 CTAGGGATTGCTGAGAGCGCTCCGAGGCGATCGCTATAGAGACCGCCAGCCAGGCGAT 64
 DB 2076 CAACATCACTACAGAGATGCTCTACACGATATCACTAACAGAGAGCGCTGTCCAGCGCAT 2135
 OY 65 CGCTAATGAGACACACCCAGTGCATGCCAAGAGAGAGCGCCAGGCGATCGC-- 121
 DB 2136 CGCTAAGAGAGCGCCGACCAAGAGGATCGCCAAAGAGAGAGCGCCAGGCGATCGCCAA 2195
 OY 122 CGAGAGAGCGGATCCGAGGATCGCCAGAGAGGTTGCCAGGCGATCGCCAAATGGGGT 181
 DB 2196 CGAGAGAGCGGCGCCAGGATCGCCAGAGAGAGCGCCAGGCGATCGCCAAAGAGAGT 2255
 OY 182 CGCCGACAGAGGCGATCGCCAGTGCATCGCCAGAGGCGATCGCCAAATGGGGCGCGT 241
 DB 2256 CGCCGCGCCAGGCGATCGCCAGAGAGAGCGCCAGGCGATCGCCAAAGAGAGAGCGCGC 2315
 OY 242 CCAGGCGTTGCGCAAGAGGAGCGCGCTCTCGCTTCCGCAAGAGAGCGCGCCAGAGG 301
 DB 2316 CCAGGCGATCGCCAGAGAGATGCTCTCGCTTCCGCAAGAGAGCGCGCGCGCAGG 2375
 OY 302 CATGCGCAAGAGGAGCGCCAGAGGCGATGGGCAAGAGGCGATCGCCAAATGGGGT 361
 DB 2376 CATGCGTAATGAGAGATCGCTATACGACATCGCTAATAGAGATCGCTAATATGACATCGC 2435
 OY 362 TACAGAGAGCGCGCTCCAGGCGATCGCTAACAGAGTGGCGCCCGAGAGCGATCGCCAAAG 421
 DB 2436 TAC--ATGACACCGCTACAGAGCA-CACGTACAGAGAGCGCTGTACACGACATCGCTAATGA 2491
 OY 422 GGAGCGCGCCAGAGGATCG--CCAGAGATGTGACAGAGGCGATGGGCAAGAGAGCG 478
 DB 2492 GGAGCGCGATCAAGAGATGGTATGAGAGAGCTGTATATGACATCGCTAACAGAGAGCAG 2551
 OY 479 CGCCGAGGCGATCGCCAAAG 500
 DB 2552 TCTACAGAGCGGTGCTAACAG 2573

RESULT 11

AA572321
 ID AAS72321 standard; cDNA; 3522 BP.

AA572321:
 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8125.
 XX
 XX
 XX

Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

OS
 XX
 PN W0200175067-A2.

PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 PI Dmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 DR P-PDB: ABG08134.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

Claim 1: SEQ ID No 8125; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 3522 BP; 1005 A; 1006 C; 959 G; 552 T; 0 other;

Query Match 50.0%; Score 250; DB 23; Length 3522;
 Best Local Similarity 74.1%; Pred. No. 5, 6e-30;

Matches 372; Conservative 0; Mismatches 120; Indels 10; Gaps 4;

OY 5 CTAGGGATTGCTGAGAGCGCTCCGAGGCGATCGCTAATGAGAGCGCCAGCGAGCGAT 64
 DB 2253 CAACATCACTACAGAGATGCTCTACACGATATCACTAACAGAGAGCGCTGTCCAGCGCAT 2132
 OY 65 CGCTAATGAGACACACCCAGTGCATGCCAAGAGAGAGCGCCAGGCGATCGC-- 121
 DB 2313 CGCTAAGAGAGCGCCGACCAAGAGGATCGCCAAAGAGAGAGCGCCAGGCGATCGCCAA 2372
 OY 122 CGAGAGAGCGGATCCGAGGATCGCCAGAGAGGTTGCCAGGCGATCGCCAAATGGGGT 181
 DB 2373 CGAGAGAGCGGCGCCAGGATCGCCAGAGAGAGCGCCAGGCGATCGCCAAAGAGT 2432
 OY 182 CGCCGACAGAGGCGATCGCCAGTGCATCGCCAGAGGCGATCGCCAAATGGGGCGCGT 241
 DB 2433 CGCCGCGCCAGGCGATCGCCAGAGAGAGCGCCAGGCGATCGCCAAAGAGAGAGCGCGC 2492
 OY 242 CCAGGCGTTGCGCAAGAGGAGCGCGCTCTCGCTTCCGCAAGAGAGCGCGCCAGAGG 301
 DB 2493 CCAGGCGATCGCCAGAGAGATGCTCTCGCTTCCGCAAGAGAGCGCGCGCGCAGG 2552
 OY 302 CATGCGCAAGAGGAGCGCCAGAGGCGATGGGCAAGAGGCGATCGCCAAATGGGGT 361
 DB 2553 CATGCGTAATGAGAGATCGCTATACGACATCGCTAATGAGAGATCGCTAATATGACATCGC 2612

PR 26-MAY-2000: 2000US-0207456.
PR 30-JUN-2000: 2000US-0608408.
PR 03-AUG-2000: 2000US-0632366.
PR 21-SEP-2000: 2000US-0234687.
PR 27-SEP-2000: 2000US-0236359.
PR 04-OCT-2000: 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25: SEQ ID NO 2635; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 495 BP; 121 A; 174 C; 129 G; 71 T; 0 other;

Query Match 47.0%; Score 235; DB 22; Length 495;
Best Local Similarity 77.3%; Pred. No. 1,3e-27;
Matches 299; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 43 ATGAGAGCGCCGACCGGATCGCTAATGAGACACCCAGATGCTGCGCAAGAGG 102
DB 108 ATGAGAGACCGGTACACATCACTAAGAGAGATGCTTCAAGATATCACTAACGAGG 167
QY 103 AAGCCGCCGAGGCGATCGC--CGAGGACGCCATCCAGAGGATCGCCACGAGAGGTTG 159
DB 168 ACGCTGTCCAGCGCATCGCTAACGAGGCGCGCCGCGCATCGCCACAGAGAGCGCG 227
QY 160 CCCAGGGCATCGGCAATGAGGCGCGGACAGAGGCGCATGAGAGAGCGCCAGG 219
DB 228 CCCAGGGCATCGGCAAGAGAGCGCGCCGCGCATCGCGGAGAGAGCGCGCCAGG 287
QY 220 GCATCGCACTGGGAGCGCGGTCCAGCGGCTTCGCCAAGGAGCGCGGTCTCAAGGTTG 279
DB 288 GCATCGCAAGAGTCCGCGCGCGCATGCGCCAGAGAGAGCGCGCCAGAGGATCG 347
QY 280 CCAAGGGGAGCGCGCCAGAGGCGATGCCAAGCGGGAGCGCCACCAAGGCGATGGGCAAG 339
DB 348 CCAAGGAGGAGCGCGGTCCAGTGGCATGCCAAGGAGAGCGGTCCAGGCGATGCTAACG 407
QY 340 AGGTACCATTCACAGGCGATGCTAACGAGAGCGCGGTCCAGGCGATGCTAACGAGGTG 399
DB 408 AGATGCGCGTGCACAGGCGATGCTAATGAGAGACTCGTATACGACATCGCTAATGAGGATG 467
QY 400 CCGGCCAGGCGATCGCCCAAGGAGAGG 426
DB 468 CCATATATGACATCGCTAATGAGACCG 494

Search completed: October 10, 2002, 15:17:57
Job time : 94 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 18.7333 Seconds
(without alignments)
6556.054 Million cell updates/sec

Title: US-09-489-101A-14_COPY_1_500

Perfect score: 500
Sequence: 1 ttagagctgctgctgctgctc.....ctaccgcgagcgccctcc 500

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	1871	1	US-08-416-478A-1
2	500	100.0	1871	2	US-08-474-988B-1
3	500	100.0	1871	2	US-08-394-442B-1
4	291	58.2	1164	1	US-08-416-478A-5
5	291	58.2	1164	2	US-08-474-988B-5
6	221	44.2	999	1	US-08-394-442B-5
7	221	44.2	999	2	US-08-416-478A-4
8	221	44.2	999	2	US-08-474-988B-4
9	221	44.2	999	2	US-08-394-442B-4
10	78	15.6	718	1	US-08-232-463-14
11	49.2	9.8	53526	3	US-08-658-136-2
12	49.2	9.8	53526	3	US-08-658-136-1
13	49.2	9.8	53526	3	US-08-658-136-1
14	47.8	9.6	289	4	US-09-007-005-17
15	47.8	9.6	289	4	US-09-244-796-17
16	42.6	8.5	320	4	US-09-165-264-11
17	40.8	8.3	4488	4	US-08-406-030A-3
18	40.8	8.2	5935	4	US-09-178-973B-17
19	40.8	8.2	5935	4	US-09-354-243B-39
20	39.6	8.1	320	4	US-09-165-264-7
21	39.6	7.9	320	4	US-09-165-264-18
22	39.2	7.8	320	4	US-09-165-264-13
23	39.2	7.8	1954	4	US-09-463-702A-32
24	39.2	7.8	3850	4	US-08-323-443B-1
25	39.2	7.8	31571	1	US-08-323-443B-1
26	39	7.8	1276	4	US-09-177-325-2
27	39	7.8	1276	4	US-09-411-812A-2

c	28	39	7.8	1276	4	US-09-590-113-2
c	29	38.8	7.8	43795	3	US-08-742-185-101
c	30	38.4	7.7	4897	6	5196516-7
c	31	37.8	7.6	318	4	US-09-165-264-12
c	32	37.6	7.5	1456	3	US-09-308-406-1
c	33	37.4	7.5	1721	5	PCR-US96-00994-3
c	34	37.2	7.4	152331	3	US-09-128-155-16
c	35	37	7.4	325	2	US-08-332-766A-11
c	36	37	7.4	12001	1	US-08-458-568A-11
c	37	36.8	7.4	283	5	PCR-US94-04361-19
c	38	36.8	7.4	3408	3	US-09-058-489-13
c	39	36.6	7.3	5322	3	US-09-058-489-13
c	40	36.6	7.3	538	2	US-08-332-766A-24
c	41	36.4	7.3	1773	4	US-08-943-711-215
c	42	36.4	7.3	6246	4	US-08-943-711-640
c	43	36.4	7.3	80246	4	US-09-078-294-4
c	44	36.4	7.3	80595	4	US-09-078-294-3
c	45	36.2	7.2	2580	3	US-09-050-863-2

ALIGNMENTS

RESULT 1
US-08-416-478A-1
Sequence 1, Application US/08416478A
Patent No. 573578
GENERAL INFORMATION:
APPLICANT: Herkend, Thierry
TITLE OF INVENTION: New Proteins Produced By Human
TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEWMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/416,478A
FILING DATE: 04-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,644
FILING DATE: 08-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: HERKEND-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1871 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 231..1724
US-08-416-478A-1
Query Match: 100.0%; Score 500; DB 1; length 1871;
Best Local Similarity: 100.0%; Pred. No. 6.2e-111;

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,442B
FILING DATE: 24-FEB-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: HERCEND-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1871 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 231..1724
US-08-394-442B-1

Query Match 100.0%; Score 500; DB 2; Length 1871;
Best Local Similarity 100.0%; Pred. No. 6, 2e-111;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCAGGTCGCTGATTCGCCAGTTTCCAGTTTCCTGATTCGAGCTCGGCTGCTGATC 60
DB 1 TCAGGTCGCTGATTCGCCAGTTTCCAGTTTCCTGATTCGAGCTCGGCTGCTGATC 60
QY 61 CTTGCCACCCCTCTCTCCAGAGGCTCTCTGCTGCTGCTCTCTCTCTCTCTCTCTCT 120
DB 61 CTTGCCACCCCTCTCTCCAGAGGCTCTCTGCTGCTGCTCTCTCTCTCTCTCTCTCT 120
QY 121 CACCTCCCTCTCTGAGAACTTCTCTTACCCGCCACCCGCCACCCGCCACTGCTTTTC 180
DB 121 CACCTCCCTCTCTGAGAACTTCTCTTACCCGCCACCCGCCACCCGCCACTGCTTTTC 180
QY 181 TTTTCTGACCTCTTTTGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 240
DB 181 TTTTCTGACCTCTTTTGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 240
QY 241 CTCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
DB 241 CTCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 AGCCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 301 AGCCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 361 GCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 361 GCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 421 AGCATGACGACGAGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
DB 421 AGCATGACGACGAGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 481 CTCACCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 500
DB 481 CTCACCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 500

RESULT 4

US-08-416-478A-5
Sequence 5, Application US/08416478A
Patent No. 5773578
GENERAL INFORMATION:
APPLICANT: Hercend, Thierry
TITLE OF INVENTION: New Proteins Produced By Human Lymphocytes, DNA Sequences Encoding These Proteins And Their Pharmaceutical And Biological Uses
TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/416,478A
FILING DATE: 04-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,644
FILING DATE: 08-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: HERCEND-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1164 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 22..1161
US-08-416-478A-5

Query Match 58.2%; Score 291; DB 1; Length 1164;
Best Local Similarity 100.0%; Pred. No. 4, 5e-61;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 210 GCTGCCAGACCTAGAGAGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 269
DB 1 GCTGCCAGACCTAGAGAGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 269
QY 270 CCGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 329
DB 270 CCGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 329
QY 330 GCCAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 389
DB 330 GCCAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 389
QY 390 AGCTTCTGAGAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 449
DB 390 AGCTTCTGAGAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 449
QY 450 GCCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 500
DB 450 GCCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 500

RESULT 5

US-08-474-988B-5
: Sequence 5, Application US/08474988B
: Patent No. 5874250
: GENERAL INFORMATION:
: APPLICANT: Hercend, Thierry
: TITLE OF INVENTION: NEW PROTEINS PRODUCED BY HUMAN
: Lymphocytes, DNA SEQUENCES ENCODING THESE PROTEINS AND
: TITLE OF INVENTION: THEIR PHARMACEUTICAL AND BIOLOGICAL USES
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
: STREET: 419 Seventh Street, N.W., Suite 400
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,988B
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/416,478
: FILING DATE: 04-APR-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/854,644
: FILING DATE: 08-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: ION, Allen C.
: REGISTRATION NUMBER: 37,971
: REFERENCE/DOCKET NUMBER: HERCEND-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 628-5197
: TELEFAX: (202) 737-3528
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 22..1161
: US-08-474-988B-5

Query Match 58.2% Score 291; DB 2: Length 1164;
Best Local Similarity 100.0%; Pred. No. 4.5e-61;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GCTGCCAGACCATAGAGAGATGTGGAGGCTCAGTTCTGTGGCTGTGTTCTGCAG 269
DB 1 GCTGCCAGACCATAGAGAGATGTGGAGGCTCAGTTCTGTGGCTGTGTTCTGCAG 60
QY 270 CCGCTTGGGTGCTCCAGTGAAGCTCTCCAGCCAGGGGCTAGAGTCCCGGTGTGTG 329
DB 61 CCGCTTGGGTGCTCCAGTGAAGCTCTCCAGCCAGGGGCTAGAGTCCCGGTGTGTG 120
QY 330 GCCCAGAGAGGGGCTCTCCAGCTCCCTGAGCCCAATATCCCTCCAGATCTC 389
DB 121 GCCCAGAGAGGGGCTCTCCAGCTCCCTGAGCCCAATATCCCTCCAGATCTC 180
QY 390 AGCTTCTGGAAGAGAGAGGGGTCACTTGGAGCATAGCCAGAGTGGCCGCGCT 449
DB 181 AGCTTCTGGAAGAGAGAGGGGTCACTTGGAGCATAGCCAGAGTGGCCGCGCT 240
QY 450 GCGGCCCGGCGCATCCCTGCGCCCGGCGCTACCGCGCGCGCGCTCC 500
DB 241 GCGGCCCGGCGCATCCCTGCGCCCGGCGCTACCGCGCGCGCGCTCC 291

DB 241 GCGGCCCGGCGCATCCCTGCGCCCGGCGCTACCGCGCGCGCGCTCC 291

RESULT 6
US-08-394-442B-5
: Sequence 5, Application US/08394442B
: Patent No. 5976877
: GENERAL INFORMATION:
: APPLICANT: Hercend, Thierry
: TITLE OF INVENTION: New Proteins Produced By Human
: Lymphocytes, DNA Sequences Encoding These Proteins And
: TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
: STREET: 419 Seventh Street, N.W., Suite 400
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/394,442B
: FILING DATE: 24-FEB-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: HERCEND-2A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 628-5197
: TELEFAX: (202) 737-3528
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 22..1161
: US-08-394-442B-5

Query Match 58.2% Score 291; DB 2: Length 1164;
Best Local Similarity 100.0%; Pred. No. 4.5e-61;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GCTGCCAGACCATAGAGAGATGTGGAGGCTCAGTTCTGTGGCTGTGTTCTGCAG 269
DB 1 GCTGCCAGACCATAGAGAGATGTGGAGGCTCAGTTCTGTGGCTGTGTTCTGCAG 60
QY 270 CCGCTTGGGTGCTCCAGTGAAGCTCTCCAGCCAGGGGCTAGAGTCCCGGTGTGTG 329
DB 61 CCGCTTGGGTGCTCCAGTGAAGCTCTCCAGCCAGGGGCTAGAGTCCCGGTGTGTG 120
QY 330 GCCCAGAGAGGGGCTCTCCAGCTCCCTGAGCCCAATATCCCTCCAGATCTC 389
DB 121 GCCCAGAGAGGGGCTCTCCAGCTCCCTGAGCCCAATATCCCTCCAGATCTC 180
QY 390 AGCTTCTGGAAGAGAGAGGGGTCACTTGGAGCATAGCCAGAGTGGCCGCGCT 449
DB 181 AGCTTCTGGAAGAGAGAGGGGTCACTTGGAGCATAGCCAGAGTGGCCGCGCT 240
QY 450 GCGGCCCGGCGCATCCCTGCGCCCGGCGCTACCGCGCGCGCGCTCC 500
DB 241 GCGGCCCGGCGCATCCCTGCGCCCGGCGCTACCGCGCGCGCGCTCC 291

RESULT 7

US-08-416-478A-4

; Sequence 4, Application US/08416478A
; Patent No. 573578

GENERAL INFORMATION:

; APPLICANT: Hercend, Thierry
; APPLICANT: Triebel, Frederic
; TITLE OF INVENTION: New Proteins Produced By Human
; TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
; TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 400
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/416, 478A; FILING DATE: 04-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/854, 644; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.; REGISTRATION NUMBER: 37, 971
; REFERENCE/DOCKET NUMBER: HERCEND-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 999 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-416-478A-4

Query Match 44.2%; Score 221; DB 1; Length 999;

Best Local Similarity 99.6%; Pred. No. 2.3e-44;

Matches 232; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 1 TCAGGCGCGTGAATGCGCCAGCTTTCAGGCTTCTCGATTCGGGCTGTCATC 60
   |||||||
DB 768 TCAGGCGCGTGAATGCGCCAGCTTTCAGGCTTCTCGATTCGGGCTGTCATC 60
   |||||||
QY 61 CTTGCCACCGCTCTCTCAAGGCGCTCTCTGTCCTCTCTTCTAGAACCCCTTC 120
   |||||||
DB 828 CTTGCCACCGCTCTCTCAAGGCGCTCTCTGTCCTCTCTTCTAGAACCCCTTC 120
   |||||||
QY 121 CACCTCCCTCTCTGCAAGAACTTCTTACCCGCCACCCACCACTGCCCCCTTTC 180
   |||||||
DB 888 CACCTCCCTCTCTGCAAGAACTTCTTACCCGCCACCCACCACTGCCCCCTTTC 180
   |||||||
QY 181 TTTTGTACCTCTTTTGGAGGGCTCAGGCGTCCAGACCATAGAGAGATG 233
   |||||||
DB 947 TTTTGTACCTCTTTTGGAGGGCTCAGGCGTCCAGACCATAGAGAGATG 233
   |||||||
```

RESULT 8

US-08-474-988B-4

; Sequence 4, Application US/08474988B

GENERAL INFORMATION:

; APPLICANT: Hercend, Thierry
; APPLICANT: Triebel, Frederic; TITLE OF INVENTION: NEW PROTEINS PRODUCED BY HUMAN
; TITLE OF INVENTION: LYMPHOCYTES, DNA SEQUENCES ENCODING THESE PROTEINS AND
; TITLE OF INVENTION: THEIR PHARMACEUTICAL AND BIOLOGICAL USES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 400
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474, 988B; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416, 478; FILING DATE: 04-APR-1995
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/854, 644; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.; REGISTRATION NUMBER: 37, 971
; REFERENCE/DOCKET NUMBER: HERCEND-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 999 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-474-988B-4

Query Match 44.2%; Score 221; DB 2; Length 999;

Best Local Similarity 99.6%; Pred. No. 2.3e-44;

Matches 232; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 1 TCAGGCGCGTGAATGCGCCAGCTTTCAGGCTTCTCGATTCGGGCTGTCATC 60
   |||||||
DB 768 TCAGGCGCGTGAATGCGCCAGCTTTCAGGCTTCTCGATTCGGGCTGTCATC 60
   |||||||
QY 61 CTTGCCACCGCTCTCTCAAGGCGCTCTCTGTCCTCTCTTCTAGAACCCCTTC 120
   |||||||
DB 828 CTTGCCACCGCTCTCTCAAGGCGCTCTCTGTCCTCTCTTCTAGAACCCCTTC 120
   |||||||
QY 121 CACCTCCCTCTCTGCAAGAACTTCTTACCCGCCACCCACCACTGCCCCCTTTC 180
   |||||||
DB 888 CACCTCCCTCTCTGCAAGAACTTCTTACCCGCCACCCACCACTGCCCCCTTTC 180
   |||||||
QY 181 TTTTGTACCTCTTTTGGAGGGCTCAGGCGTCCAGACCATAGAGAGATG 233
   |||||||
DB 947 TTTTGTACCTCTTTTGGAGGGCTCAGGCGTCCAGACCATAGAGAGATG 233
   |||||||
```

RESULT 9

US-08-394-442B-4

; Sequence 4, Application US/08394442B

GENERAL INFORMATION:

; APPLICANT: Hercend, Thierry
; APPLICANT: Triebel, Frederic
; TITLE OF INVENTION: New Proteins Produced By Human
; TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
; TITLE OF INVENTION: Their Pharmaceutical And Biological Uses

```

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESS: BROMDY AND NEWMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,442B
FILING DATE: 24-FEB-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BROMDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: HERCEND-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-394-442B-4

Query Match
Best Local Similarity 44.2%; Score 221; DB 2; Length 999;
Matches 232; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 TCAGGCTGCTGATGTCGCCAGCTTTCAGCTTCTCTGATTCGGGCTCTGTCATC 60
DB 768 TCAGGCTGCTGATGTCGCCAGCTTTCAGCTTCTCTGATTCGGGCTCTGTCATC 827
QY 61 CTCGCCAGCTTCTCTGAGGCGCTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCT 120
DB 828 CTCGCCAGCTTCTCTGAGGCGCTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCT 887
QY 121 CACGCTGCTCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 888 CACGCTGCTCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 946
QY 181 TTTCTGACCTCTTTGGAGGCTCAGGCTGCGCCAGACCATAGAGAGATG 233
DB 947 TTTCTGACCTCTTTGGAGGCTCAGGCTGCGCCAGACCATAGAGAGATG 999

RESULT 10
US-08-232-463-14
Sequence 14, Application US/08232463
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)856-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-F15
US-08-232-463-14

Query Match
Best Local Similarity 15.6%; Score 78; DB 1; Length 7218;
Matches 12; Conservative 142; Mismatches 132; Indels 0; Gaps 0;

QY 12 GATCTGCCAGCTTTCAGCTTCTCTGATTCGGGCTCTGTCATCCTCCGCCACC 71
DB 1064 GATCTGCCAGCTTTCAGCTTCTCTGATTCGGGCTCTGTCATCCTCCGCCACC 1123
QY 72 TCCTCCAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 131
DB 1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1183
QY 132 CTCGAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 191
DB 1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1243
QY 192 CTTTGGAGGCTCTGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 251
DB 1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1303
QY 252 GCTTGTCTGTCGAGCGCTTGGGTGCTCCAGTGAAGCTCTCCAGCCAGGGCT 311
DB 1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1363
QY 312 GAGTCCGCGTGTGAGGCGCCAGAGAGGCGCTCTCTCTCTCTCTCTCTCTCT 371
DB 1364 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1423
QY 372 ATCCCTCCAGAGCTCAGCTTCT 397
DB 1424 YYYYYYYYYYGTACCAATCTCTCT 1449

RESULT 11
US-08-658-136-2
Sequence 2, Application US/08658136
GENERAL INFORMATION:
APPLICANT: KLINER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY

```

```

: APPLICANT: QIAN, FENG
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: ONE MOUNTAIN ROAD
: CITY: FRAMINGHAM
: STATE: MASSACHUSETTS
: COUNTRY: USA
: ZIP: 01701
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/658,136
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: LASSEN, ELIZABETH
: REGISTRATION NUMBER: 31,845
: REFERENCE/DOCKET NUMBER: GEN4-17.8
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 508-872-8400
: TELEFAX: 508-872-5415
:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 53526 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-658-136-2
:
: Query Match          9.8%; Score 49.2; DB 3; Length 53526;
: Best Local Similarity 55.2%; Pred. No. 0.0065;
: Matches 96; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
:
: QY 50 CTCGTGTCATCCCTCCACCCCTCTCTCCAGGCCCTCTCTGAGTCTCCCTTCTCTAGA 109
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 36459 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 36518
:
: QY 110 ACCCTTCTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 169
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 36519 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 36578
:
: QY 170 GCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 223
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 36579 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 36632
:
: RESULT 12
: US-08-658-136-1
: Sequence 1, Application US/08658136
: Patent No. 6071717
: GENERAL INFORMATION:
: APPLICANT: KLINGER, KATHERINE W
: APPLICANT: LUNDEG, GREGORY M
: APPLICANT: BURN, TIMOTHY C
: APPLICANT: CONNORS, TIMOTHY D
: APPLICANT: DACKOWSKI, WILLIAM
: APPLICANT: GERMINO, GREGORY
: APPLICANT: QIAN, FENG
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: ONE MOUNTAIN ROAD
: CITY: FRAMINGHAM
: STATE: MASSACHUSETTS
: COUNTRY: USA
: ZIP: 01701

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/658,136
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: LASSEN, ELIZABETH
: REGISTRATION NUMBER: 31,845
: REFERENCE/DOCKET NUMBER: GEN4-17.8
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 508-872-8400
: TELEFAX: 508-872-5415
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 53577 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-658-136-1
:
: Query Match          9.8%; Score 49.2; DB 3; Length 53577;
: Best Local Similarity 55.2%; Pred. No. 0.0065;
: Matches 96; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
:
: QY 50 CTCGTGTCATCCCTCCACCCCTCTCTCCAGGCCCTCTCTGAGTCTCCCTTCTCTAGA 109
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 36454 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 36513
:
: QY 110 ACCCTTCTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 169
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 36514 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 36573
:
: QY 170 GCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 223
:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: Db 36574 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 36627
:
: RESULT 13
: US-09-007-005-17/C
: Sequence 17, Application US/09007005B
: Patent No. 6238358
: GENERAL INFORMATION:
: APPLICANT: Szostak, Jack W.
: APPLICANT: Roberts, Richard W.
: APPLICANT: Liu, Rihue
: TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
: FILE REFERENCE: 00786/350003
: CURRENT APPLICATION NUMBER: US/09/007,005B
: EARLIER FILING DATE: 1998-01-14
: EARLIER APPLICATION NUMBER: 60/035,963
: EARLIER FILING DATE: 1997-01-27
: EARLIER APPLICATION NUMBER: 60/064,491
: EARLIER FILING DATE: 1997-11-06
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 17
: LENGTH: 289
: TYPE: RNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Translation template
: NAME/KEY: misc.feature
: LOCATION: (1)..(289)
: OTHER INFORMATION: n = A,T,C or G
: US-09-007-005-17

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds
(without alignments)

9699.805 Million cell updates/sec

Title: US-09-489-101A-14_COPY_1-500
Perfect score: 500

Sequence: 1 tcagtgctgcctgatctgcc.....ctcacccggcgcgccctcc 500

Scoring table: IDENTITY_NUC
GapPen 10.0 Gapext 1.0

Searched: 13736207 seas. 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em estba: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	298	59.6	586	10	BC989429	BG989429 HOA55.1-1
2	232	46.4	377	10	BC989484	BG989484 HOA19.1-1
3	227	45.4	381	10	BC980314	BG980314 HOA37.1-1
4	150.8	30.2	602	10	BI060106	BI060106 IL3-UTR0111
5	149.2	29.8	408	10	BI060032	BI060032 IL3-UTR0111
6	68.8	12.6	538	9	AL514267	AL514267 AL514267
7	63	12.6	764	9	AL547921	AL547921 AL547921
8	61.8	12.4	960	12	AG060752	AG060752 Pan trogl
9	60.8	12.2	1222	10	BI166600	BI166600 haasp001xrg
10	59.6	11.9	987	12	CNS00418	AL066507 Drosophi11
11	59	11.8	917	12	AG081214	AG081214 Pan trogl
12	58.6	11.7	796	12	AG10286	AG10286 Pan trogl
13	58.6	11.7	945	12	CNS0076K	AL066688 Drosophi11
14	57.8	11.7	948	12	CNS0091B	AL074439 Drosophi11
15	57.8	11.6	925	12	CNS017P	AL033013 Drosophi11
16	57.2	11.4	101	12	CNS0147P	AL08415 Drosophi11
17	56.6	11.3	926	12	AG061568	AG061568 Pan trogl

C	18	56.6	11.3	997	12	CNS00505E
C	19	56.6	11.2	932	12	AG134650
C	20	56.2	11.2	949	12	AG078085
C	21	55.8	11.2	966	12	CNS02NRG
C	22	55.8	11.2	1042	12	BI416535
C	23	55.8	11.2	1101	12	CNS015523
C	24	55.6	11.1	1183	12	BG849405
C	25	55.6	11.1	1183	12	CNS0166R
C	26	55.4	11.1	1141	10	BF131888
C	27	55.4	11.1	1236	12	AG126182
C	28	55.2	11.0	703	12	AG194432
C	29	55.2	11.0	1451	10	BF219446
C	30	55	11.0	724	12	AG117201
C	31	54.8	11.0	959	12	CNS008842
C	32	54.6	10.9	883	10	BI887282
C	33	54.4	10.9	1201	12	CNS015XQ
C	34	54.2	10.8	788	12	AG032933
C	35	54.2	10.8	901	12	AG043735
C	36	54.2	10.8	1110	10	BM479447
C	37	54	10.8	663	12	AG142296
C	38	53.8	10.8	768	12	AG1610225
C	39	53.8	10.8	419	9	A1268505
C	40	53.8	10.8	672	12	BH140198
C	41	53.8	10.8	689	12	CNS03XPX
C	42	53.8	10.8	776	12	AG081619
C	43	53.8	10.8	796	12	AG143695
C	44	53.8	10.8	952	10	BG433836
C	45	53.8	10.8	1061	12	AG072444

ALIGNMENTS

RESULT 1				
LOCUS	BG898429			
DEFINITION	BG898429	586 bp	mrna	linear
	HOA55-1-F8.R HOA (Human Osteoarthritis Cartilage)			Homo sapiens cDNA

ACCESSION BG898429 , mRNA sequence.

VERSION BG898429.1 GI:14308678
KEYWORDS EST.

SOURCE ORGANISM	human.
	Homo sapiens

.....

REFERENCE
AUTHORS

344

3711

JOURNAL

MEDLINE
COMMENT

FEATURES

TNOs

PAGE CONTINUED

BASE COIN
ORIGIN

Query Match	59.6%	Score 298	DB 10	Length 586
Best Local Similarity	100.0%	Pred. No. 3,6e-47		
Matches 298	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	203	GGTCACCGCTGCCCGACCATAGAGAGATGTGGAGAGCTCACTTCCTGGCTTGCTGT	262	
Db	83	GGTCACCGCTGCCCGACCATAGAGAGATGTGGAGAGCTCACTTCCTGGCTTGCTGT	142	
Oy	263	TTGCGACCGCTTTGGGAGGCTCCATATGAAGCTCTCCACCGAGGGCTGAGAGTCCGGT	332	
Db	143	TTGCGACCGCTTTGGGAGGCTCCATATGAAGCTCTCCACCGAGGGCTGAGAGTCCGGT	202	
Oy	323	GGTGTGGGCGCAGAGAGGGGCTCCTGCGCAGCTCCCTCGAGGCGCCAGCAATCCCTCCCA	362	
Db	203	GGTGTGGGCGCAGAGAGGGGCTCCTGCGCAGCTCCCTCGAGGCGCCAGCAATCCCTCCCA	262	
Oy	383	GGATCTACACCTTCTGGAGAGACAGGGGTACTGTGGACATACACCGACATGTGGCC	442	
Db	263	GGATCTACACCTTCTGGAGAGACAGGGGTACTGTGGAGATACACAGAAATGGGGCC	322	
Oy	443	GCCCCGTGCCCCCCCCCGCCATCCCTGAGCCCCCGCCCTTCACCGCGGCGCCCTCC	500	
Db	323	GCCCCGTGCCCCCCCCCGCCATCCCTGAGCCCCCGCCCTTCACCGCGGCGCCCTCC	380	
RESULT 2				
EG898484	377 bp	mrna	linear	EST 06-NOV-2001
LOCUS	HOA19-1-A4 HOA (Human Osteoarthritic Cartilage)	Homo sapiens cDNA,		
DEFINITION	mrna sequence.			
ACCESSION	EG898484			
VERSION	EG898484.1	GI:14308733		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 377)			
JOURNAL	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,			
MEDLINE	Lark,M.W., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M., and			
COMMENT	21482651			
	Identification and initial characterization of 5000 expressed			
	sequenced tags (ESTs) each from adult human normal and			
	osteoarthritic cartilage cDNA libraries			
	Osteoarthrits Cartilage 9 (7), 641-653 (2001)			
	Contact: Sanjay Kumar			
	UM2109			
	GlaxoSmithKline			
	709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA			
	Tel: 610-270-7245			
	Fax: 610-270-5598			
	Email: sanjay.kumar-1@sk.com			
	Seq primer: T7.			
FEATURES	Location/Qualifiers			
source	1..377			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"			
	/tissue_type="Cartilage"			
	/lab_host="E.coli DH10 B"			
	/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;			
	Directional"			
BASE COUNT	53 a 163 c 71 g 89 t 1 others			
INSD				
Query Match	46.4%	Score 232	DB 10	Length 377
Best Local Similarity	99.6%	Pred. No. 1.2e-34		
Matches 232	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Oy	1	TGAGGCTCGCGATGAGCCGACGATTTCCAGCTTCTCTGATTCGGCCCTGTGATC	60	

Dd	145	TGAGGGTCGTGAATCTGCCACGATTTCCAGCTTTCCTCCTCGATGCATTCCGGCTCTGTGTCATC	204					
Oy	61	CTTCCCACACCCTCTCTCCAAAGGCCCCCTCTCTGTCGTCCCTCTTTCTTAGAACCCCTTCCTC	120					
Dd	205	CCCTCCACACCCCTCTCTCCAAGGCCCTCTCTGTGTCGTCCCTCTTTCTTAGAACCCCTTCCTC	264					
Oy	121	CACCTCCCTCTCTCAAGAATCTCTCTTAACCCCCACCCACCACACTGCCCCCTTTTC	180					
Dd	265	CACCTCCCTCTCTCAAGAATCTCTCTTAACCCCCACCCACCACACTGCCCCCTTTTC	324					
Oy	181	TTTTTGCACTCTCTTTGGAGGGCTCAGCGCTGCCAAGACCATAGAGAGATG	233					
Dd	325	TTTTTGCACTCTCTTTGGAGGGCTCAGCGCTGCCAAGACCATAGAGAGATG	377					
RESULT 3								
Bg900314		381 bp	mRNA linear EST 06-NOV-2001					
Bg900314								
LOCUS								
DEFINITION								
ACCESSION								
VERSION								
KEYWORDS								
SOURCE								
ORGANISM								
Human.								
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
REFERENCE								
AUTHORS								
Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J.,								
Sathe,G.V., Mui,P., Agrawal,P., Badger,A.M., Lee,J.C., Gowen,M. and								
Lark,M.W.								
Identification and initial characterization of 5000 expressed								
sequenced tags (ESTs) each from adult human normal and								
osteoarthritis cartilage cDNA libraries								
Osteoarthritis Cartilage 9 (7), 641-653 (2001)								
TITLE								
JOURNAL								
MEDLINE								
COMMENT								
Contact: Sanjay Kumar								
UM2109								
GlaxoSmithKline,								
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA								
Tel.: 610-270-7745								
Fax: 610-270-5598								
Email: sanjay.kumar.1@sk.com								
Seq primer: f7								
Location/Qualifiers								
source								
1								
organism="Homo sapiens"								
db_xref="taxon:9606"								
clone_lib="HOA (Human Osteoarthritis Cartilage)"								
tissue_type="Cartilage"								
lab_host="E.coli DH10 B"								
note="Vector: pSPORT I; Site_1: Sall; Site_2: NotI;								
Directional"								
BASE COUNT								
ORIGIN								
51 a 165 c 70 g 94 t 1 others								
Query Match								
Best Local Similarity 45.4%; Score 227; DB 10; Length 381;								
Matches 227; Conservative 0; Pred.No.Ie=35; Indels 0; Gaps 0;								
Oy	1	TCAGAGCTGCATATCGCCACAGCTTTCACAGCTTCTCTCTGATTCGGCCCTGTGATATC	60					
Dd	154	TCAGAGTGTGATATCGCCAGAGCTTTCACAGCTTCTCTCTGATTCGGCCCTGTGATATC	213					
Oy	61	CTTCCCACACCCTCTCTCAAAGGCCCTCTCTGTGTCGTCCCTCTTTCTTAGAACCCCTTCCTC	120					
Dd	214	CTTCCCACACCCTCTCTCAAAGGCCCTCTCTGTGTCGTCCCTCTTTCTTAGAACCCCTTCCTC	273					
Oy	121	GACCTCCCTCTCTCAAGAACCTCTCTTAACCCCCACCCACCACACTGCCCCCTTTTC	180					
Dd	274	GACCTCCCTCTCTCAAGAACCTCTCTTAACCCCCACCCACCACACTGCCCCCTTTTC	333					
Oy	181	TTTTTGCACTCTCTTTGGAGGGCTCAGCGCTGCCAAGACCATAGAGAGATG	228					

Db 334 TTTCTGACCTCTTTGGAGGCGTCAGCCCTGCCAGCAGCAGANAG 381
|||||
RESULT 4
LOCUS BI060106 602 bp mRNA linear EST 15-JUN-2001
DEFINITION IL3-UT014-160101-420-A06 UT014 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI060106
VERSION BI060106.1 GI:14467633
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 602)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bionesi,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?IL-IL3&t2-IL3-UT014-
160101-420-A06&t3-2001-01-16&t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 601.
Location/Qualifiers
1..602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT014"
/dev_stage="Adult"
/note="Organ: uterus; tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORFESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 104 a 228 c 149 g 121 t
ORIGIN
Query Match 30.2%; Score 150.8; DB 10; Length 602;
Best Local Similarity 95.7%; Pred. No. 3,4e-19;
Matches 155; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

RESULT 5
LOCUS BI060032 408 bp mRNA linear EST 15-JUN-2001
DEFINITION IL3-UT014-160101-419-A10 UT014 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI060032
VERSION BI060032.1 GI:14467559
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 408)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bionesi,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?IL-IL3&t2-IL3-UT014-
160101-419-A10&t3-2001-01-16&t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 377.
Location/Qualifiers
1..408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT014"
/dev_stage="Adult"
/note="Organ: uterus; tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORFESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 76 a 163 c 90 g 79 t
ORIGIN
Query Match 29.8%; Score 149.2; DB 10; Length 408;
Best Local Similarity 95.1%; Pred. No. 6,4e-19;
Matches 134; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 285 CCAGTGAAGCTCTCCAGCCAGGAGGCTGCGTGTGGCCAGAGAGGGGCT 344
|||||
Db 125 CTATGTAAGCTCTCCAGCCAGGAGGCTGCGTGTGGCCAGAGAGGGGCT 184
|||||
QY 345 CCGGCCAGCTCTCCAGCCAGGAGGCTGCGTGTGGCCAGAGATCTCAGCTTCGAGAGA 404
|||||
Db 185 CCGGCCAGCTCTCCAGCCAGGAGGCTGCGTGTGGCCAGAGATCTCAGCTTCGAGAGA 244
|||||
QY 405 GCAGGGGTCACTTGGCAGCATCAGCCAGCAGATGGCCGCC 446
|||||
Db 245 GCAGGGGTCACTTGGCAGCATCAGCCAGCAGATGGATGACCC 286
|||||

RESULT 6
AL514267 538 bp mRNA linear EST 13-FEB-2001
DEFINITION AL514267 LFL_NFL006_P12 Homo sapiens cDNA clone CL08B0072D09 3

Fri Oct 11 09:30:44 2002

us-09-489-101a-14_copy_1_500.rst

Page 4

prime, mRNA sequence.
ACCESSION AL514267
VERSION AL514267.1 GI:12777761
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 538)
AUTHORS L.W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry Cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="CD05072D06"
/issue="1"
/issue_type="Placenta"
/note="Vector: PCWSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWSPORT 6 vector. Library was normalized. Library was constructed by a division of Invitrogen 9800 Medical Center Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
http://fulllength.invitrogen.com"

BASE COUNT
3 a 299 c 103 g 85 t 48 others
ORIGIN

Query Match 14.0% Score 69.8; DB: 9; Length 538;
Best Local Similarity 42.7%; Pred. No. 0.00083;
Matches 198; Conservative 24; Mismatches 242; Indels 0; Gaps 0;
QY 32 TTTCCTGTGATTCGCGCTGTCATCCCTCCGACCTCTCTCCAGGCGCTCTCT 91
DB 75 TTTTCTTTTTCCT 134
QY 92 GGTCT 151
DB 135 CC 194
QY 152 CC 211
DB 195 CC 254
QY 212 TGGCAGACATAGAGATGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 271
DB 255 CCCCCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314
QY 272 GCTTGGGCT 331
DB 315 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 374
QY 332 CAGAGAGGCGGCT 391
DB 375 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 434
QY 392 CTTCT 451
DB 435 YGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 494
QY 452 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 495
DB 495 CCCCCCCCCCTTCT 538

RESULT 7
AL547921/c 764 bp mRNA linear EST 16-FEB-2001
LOCUS AL547921 LTI_NFL006_PL2 Homo sapiens cDNA clone CS001033YH07 3
DEFINITION prime, mRNA sequence.
ACCESSION AL547921
VERSION AL547921.1 GI:12882437
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 764)
AUTHORS L.W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry Cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="CS001033YH07"
/issue="1"
/issue_type="Placenta"
/note="Vector: PCWSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWSPORT 6 vector. Library was normalized. Library was constructed by a division of Invitrogen 9800 Medical Center Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
http://fulllength.invitrogen.com"

BASE COUNT
174 a 170 c 238 g 23 t 159 others
ORIGIN

Query Match 12.6% Score 63; DB: 9; Length 764;
Best Local Similarity 31.2%; Pred. No. 0.017;
Matches 149; Conservative 93; Mismatches 235; Indels 0; Gaps 0;
QY 5 GGTCT 64
DB 503 SCTKSCSSTCT 444
QY 65 CCGACCT 124
DB 443 STCT 384
QY 125 TCCCT 184
DB 383 TCCSCKSSTCT 324
QY 185 CTGACCT 244
DB 323 SCCCT 264
QY 245 GTTCT 304
DB 283 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 364
QY 305 AGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 364
DB 203 TGTCT 144
QY 365 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 424
DB 143 CTTCT 84
QY 425 TCGACGACATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 481

DB 83 GCTAGTCCAGGCTCCTCCTGAGCGGTCCTTCCGACGCCACGATCC 27

RESULT 8
LOCUS AG060752 960 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-048j12.F, genomic survey sequence.
ACCESSION AG060752.1 GI:16612554
VERSION AG060752.1
KEYWORDS GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
BAC library clone:PTB-048j12.F.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.
TITLE Unpublished
JOURNAL BAC end sequences of library PTB
REFERENCE 2 (bases 1 to 960)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpesgsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13

LIBRARY
Vector : pKSI45
R.Site 1 : SacI
R.Site 2 : Sact.
Location/Qualifiers
1. 960
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-048j12.F"
/sex="male"
/cell_type="lymphoblast"
/clone_11b="PTB Chimpanzee Male BAC Library"
BASE COUNT 93 a 24 c 670 g 38 t 135 others
ORIGIN

Query Match 12.4%; Score 61.8; DB 12; Length 960;
Best Local Similarity 39.3%; Pred. No. 0.031;
Matches 192; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

QY 10 CTGATGTCGAGCTTTCAGGCTTCTGATTCGGGCTCTGTCATCCGCCAC 69
DB 546 CANNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 487
QY 70 CCTCTTCAGAGGCTCTCTGTCCTCTTCTTGAAGACCTCTCTCTCTCT 129
DB 486 CCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 427
QY 130 CTCTGCAAGACTTCTCTTACCCGCCGCCGCCGCCGCCGCCGCCGCC 189
DB 426 CNCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 367
QY 190 CTCCTTTGAGGCTCTCTGTCAGGCTCTGAGAGATGTGAGAGGCTCT 249
DB 366 CCCCCCT 307
QY 250 TGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309
DB 306 NTAAACCCAGTTGGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 247

QY 310 CTGAGTCCCGCTGTGTGGGCCAGAGGGGGCTCTGACCTCCCTGAGCCCA 369
DB 246 CNNCCACAACAACCAACCAACCAACCAACCAACCAACCAACCAACCA 187
QY 370 CAAATCCCTTCAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 429
DB 186 CC 127
QY 430 CAGACATGAGGCGCCGCTCTGTCGCGCCGCCGCCGCCGCCGCCGCC 489
DB 126 CCAAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 67
QY 490 CGGCGCC 497
DB 66 NNCCGCC 59

RESULT 9
LOCUS B141660/c 1222 bp mRNA linear EST 15-AUG-2001
DEFINITION haspD01xm19f Heterobasidion annosum - Scots pine infection stage
subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone haspD01xm19f, mRNA sequence.
ACCESSION B141660
VERSION B141660.1 GI:15187683
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1222)
AUTHORS Asiegbu, F.O., Nakhkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the
infection of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
CONTACT: Fred O. Asiegbu
DEPT. OF Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
TEL: +46 18 67 15 98
FAX: +46 18 30 92 45
EMAIL: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

FEATURES
source
1. 1222
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="haspD01xm19f"
/clone_11b="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="Vector: pT-Adv; Site: 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6 days with mycelia of
Heterobasidion annosum (FP5)."
BASE COUNT 109 a 23 c 1007 g 4 t 79 others
ORIGIN

Query Match 12.2%; Score 60.8; DB 10; Length 1222;
Best Local Similarity 42.4%; Pred. No. 0.049;
Matches 206; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

QY 15 CTGCGAGCTTTCAGGCTTCTGATTCGGGCTCTGTCATCCGCCACCTCT 74
DB 490 CTTCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 431
QY 75 CTCGAAGGCTCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 134
DB 430 CCGTCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 371
QY 135 CAGAACTTCTCTTACCGCCGACGCCGCCGCCGCCGCCGCCGCCCT 134

[illegible]

RESULT 10	CNS00418	LOCUS	CNS00418	987 bp	DNA	linear	GEN 03-JUN-1999
DEFINITION	BACR09c16 of RPEC-98 library from <i>Drosophila melanogaster</i> (fruit fly); genomic survey sequence.						
Accession	U02553						

ACCESSION	AL066537
VERSION	AL066537.1
KEYWORDS	GSS.
SOURCE	Fruit fly.

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 987)

REFERENCE

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a

melanogaster BAC library was prepared by Gatzky/Osogawa and Aaron Memmoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila.bac.htm>.

FEATURES	Location/Qualifiers
source	1. .987

[illegible]

Query Match	11.9%;	Score 59.6;	DB 12;	Length 987;
Best Local Similarity	23.5%;	Pred. No. 0.08;		
Matches 92;	Conservative 123;	Mismatches 177;	Indels 0;	Gaps 0.

[illegible]

RESULT 11	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM
AG081214/c	AG081214	Pan troglodytes DNA, clone: PTB-077j19.R, genomic survey sequence.	AG081214	AG081214.1 GI:16633016	GSS: GSS (genome survey sequence). Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC library clone:PTB-077j19.R. Pan troglodytes	917 bp DNA linear GSS 03-NOV-2003

REFERENCE
1 (sites)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.

REFERENCE	AUTHORS	TITLE
1 (sites)		
	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
	Notoki, Y., Watanabe, H. and Sakaki, Y.	
	BAC end sequences of library PB8	

JOURNAL, Unpublished
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (02-0UG-2001) Asao Fujiyama, The Institute of
Direct Submission
Tokio, Y., Watanabe, H. and Sakaki, Y.
Fujiyama, A., Battori, M., Toyoda, A., Taylor, T. D., Yada, T.
2 (bases 1 to 917)

1-7-22, Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.
E-mail: chianpbes@sc.riken.go.jp, URL: <http://hnp.gsc.riken.go.jp/>,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PNB. This BAC end
was generated during the RAD process and may have higher chance of
clone tracking errors.

Sequencing: M13Rev
LIBRARY

```

Vector      : PKS145
R.Site 1   : Sac1
R.Site 2   : Sac1.
Location/Quali
FEATURES

```

```
source
1. .917
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-077J19.R"
```

BASE COUNT	ORIGIN	68 a	35 c	666 g	46 t	102 others
/sex="male"						
/cell_type="lymphoblast"						
/clone_lib="PTB Chimpanzee Male BAC Library"						

Query Match	11.8%;	Score 59;	DB 12;	Length 917;
Best Local Similarity	41.3%;	Pred. No. 0.1;		
Matches 188;	Conservative	0;	Mismatches 267;	Indels 0;
			Gaps	0;

Qy	45	CCGGCCTCTGGTATCCCTCCCAACCTCTCTCCAAAGGCGCTCTCGTGGTCCCTCTT	104
Db	553	CCCMCCCCCCCCCTCC	494
Qy	105	CTPAGAACCCCTTCCCTCCAGCTCCCTCTCTGACAGAACTTCTCTTAAACCCCAACCCCA	164
Db	493	CCCN	434
Qy	165	CCACTGGCCCCCTTCCCTTTCCTTTTCTAGACCTCTTTTGGAGGGCTCAAGCGCTGCCAGACATA	224
Db	433	CCCCCCCCCGCCCCCCCCCCCCCNCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCN	374
Qy	225	GGAGAGATGTGGAGGCTCACTTCCTGGGCTTGCTGTCTTGCAGCCGCTTGGGTGGCT	284
Db	373	CCCCCCCCCCCCCNCCCCCNCCNCCCCCCCCCNCCNCCNCCCCCCCCCCCCCCCC	314
Qy	285	CCAATGAAGCTCTCCAGGCCAGGGGGCTGAGGTCCCCGCTGTGTGGGCCCAAGAGGGGGCT	344
Db	313	NNNNNNNNCCCCCCCCCNCCNNCNAATGTCCACCCCCAGGGGGGGGTGGGGGCGCGCGC	254
Qy	345	CCCTGGCCAGCTCCCTCTGAGGCCACAAATCCCTCTCCAGATCTCAGCTTCTGCAAGA	404
Db	253	CCCGGCCANCCCCCCCCCACCACACACANCCCCACCAANNNNNNNNCANNNNNNCC	194
Qy	405	GCAGGGGTCACTTGGCAGCATCAGCCACAGACAGTGGCCCGCCGCTGCCGCCCCGGCCAT	464
Db	193	CCCCNNCCCCCCCCCGCCGCCACCCACCAAAACACCCCCCCCCCCCCCCCCCCCC	134
Qy	465	CCCCGTGGCCCCGGGCGCTCAACGGGGGGCGGCGCTC	499
Db	133	CCCCCCCCNCCCCCACAACCCCCCNCCCCCCCCCCCC	99

RESULT 12
AG110286/c

LOCUS AG110286 796 bp DNA Linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-115P11.R, genomic survey sequence.
ACCESSION AG110286
VERSION AG110286.1 GI:16730805
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
BAC library clone:PTB-115P11.R.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
unpublished
2 (bases 1 to 796)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpdesc@sc.riken.go.jp, URL:htp://npg.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS

```

Sequencing: M3Rev
LIBRARY
  Vector      : pKS145
  R.Site 1    : SacI
  R.Site 2    : SacI.
Location/Qualifiers
  1..796
     /contig="Baan_17001096"

```

	/clone_lib="PTB Chimpanzee Male BAC Library"			
BASE COUNT	59 a	59 c	644 g	10 t
ORIGIN				24 others

Query Match	11.7%	Score 58.6	DB 12	Length 796
Best Local Similarity	44.3%	Pred. No. 0.12		
Matches 220; Conservative	0	Mismatches 277	Indels 0	Gaps 0

[illegible]

RESULT 13	
LOCUS	CNS0076K
DEFINITION	CNS0076K 945 bp DNA linear GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence Tetr3 end of BAC #
VERSION	BAC114C09 of RPCI-98 library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
SOURCE	AL066880
ORGANISM	AL066880.1 GI:4945343
REFERENCE	GSS.
AUTHORS	fruit fly.
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 945)
	Genoscope.

TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segrete@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammosser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
 1. 945
SOURCE /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACRIA.P09"
 /note="end : TET3"
BASE COUNT 165 a 308 c 83 g 185 t 204 others
ORIGIN
 Query Match 11.7%; Score 58.6; DB 12; Length 945;
 Best Local Similarity 32.9%; Pred. No. 0.12;
 Matches 104; Conservative 74; Mismatches 136; Indels 0; Gaps 0
 QY 18 CCCAGCTTCCAGCTTCTCCTGTGATTCGGCGCTGTGCATCCCTCCCAACCTCTCTC 77
 ||| :
 Db 565 CCCSSCGSYCTCTCBBCCCCCCCCCCCKKCGCBTCCBCCCTCCCCCCCCCCCC 624
 QY 78 CAAGGCGCTCCTGGTGTCTCCTTTCTTGAAGACCCCTTCTCCACCTCCCTCTCTCGAG 137
 | :
 Db 625 CCKKCCCTTTCBBSBGGCCCTTBBCCCCCBBCBCCCCGCTCKKCCCCCCCCCCC 684
 QY 138 AACTCTCTCTTACCCGCCACACCCACACACGCCCCCTTCTCTTTCTTGACCTCTCTTT 197
 ||| : ||||| | ||| | |||| | : : : : : : : : : : : : : : : : : :
 Db 685 CCKCTCTCKKCKCCCCCGCKCCCCCCCCCGCCGCGTGKGYSYCBTBOBCH 744
 QY 198 GGAGGCGTCAGGCGTCCCAAGACATAGAGAGAGTGGAGGCTCAGTCTCGAGCTTG 257
 745 TCTKTKTCTTGBCTBCCSYSCBSGGBKCKSKTKGTSGTSGKGTCTSCCGKTY 804
 QY 258 CTGTTTTCGACGCGCTTGGGTGCTCAGTAGAAGCTCTCCACACAGGAGCTAGAGTC 317
 :
 Db 805 SYGTCTCKSCGVBKSTSTSCCKSKYCTGTGGCVCBCTBRYTSSGCBCKCKSKSCGS 864
 QY 318 CCGGTGGTGTGGGCC 333
 :
 Db 865 SCCKCKBKSBSGCCS 880
RESULT 14
CNS001CB
LOCUS CNS001CB 948 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
 BACR36N04 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
ACCESSION AL074499
VERSION AL074499.1 GI:4954380
SOURCE GSS.
KEYWORDS fruit fly.
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 948)

AUTHORS	TITLE	Genoscope.
JOURNAL	Direct Submission	
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefr@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC and sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammotser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCl-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.bufralo.edu/drosophila_bac.htm . Location/Qualifiers	
FEATURES	source	1..948 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RpCl-98" /clone="BACR36N04" /note="end : TET3"
BASE COUNT	130 a 355 c 125 g 146 t 192 others	
ORIGIN		
Query Match	11.7% Score 58.6: DB 12; Length 948;	
Best Local Similarity	35.4% Pred. No. 0.12; Tm 75; Gaps 1;	
Matches 154; Conservative 75; Mismatches 203; Indels 3; Gaps 1;		
QY	69 CCCTCTCTCCAGAGCGCCCTCTCGTGGTCTCCCTCTTCTAGAACCCCTTCTCCACCTCCC	128
DB	341 WMCCTGTGKCGKCGGGCGGCGCTTGTGGVTTCTYTCGGCGHMYCAGCGCCYGTGCC	400
QY	129 TCTCTGACAGACTTCTCTTTACCCCCACCCCCACACACTGCCCCCTTCTCTTCTGA	188
DB	401 GGYTGGCGCTTCCCGCCCTTCCCGCGCTTTTCCCGCCCTTCCCGCCCTTCCCGCCCTT	460
QY	189 CCTCTTTTGGAGGCTCAGCGCTGCCAGACATAGAGAGATGTGGAGGCTCAGTTTC	248
DB	461 CTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	520
QY	249 CTGGGCTTCTTCTTCTGACGCGCTTTGGGTGCTCCAGTGAAGCTCTTCCAGCCAGG	308
DB	521 TTSGGGSCCCSTVBGVSWSMTGCCCTCBBYGTAKTCTTCTTCTTCTTCTTCTTCTTCT	580
QY	309 GCTGAGGTCCGGGTGTGTGGGGGCGAGAGAGGGGGG---TCTGCGCCACTTCCCTGACG	365
DB	581 CCCCCBCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	640
QY	366 CCACATATCCCTTCCAGAGATCTGACCTTCTGCGACAGAGAGGGGTCTTGGAGCAT	425
DB	641 CCCCCGCGCTCCCGCGGGBBDBCCYCCGCCSCCCSSCCBCTGCTTCTGKTGKCGG	700
QY	426 CAGCCAGACAGTGGCCCGCGCTGCGCCCGCCGCGCATCCCTGGGCCCGGCGCTCAC	485
DB	701 GCKCTGSGGTTTGGKCGKSCCCCCCCCCSCCCSCCCSCCCSCCCSCCKGKSCSTYGS	760
QY	486 CCGGCGGGGCGGCTTC 500	
DB	761 CCTGAGCCCGKCYC 775	
RESULT 15		
LOCUS	CNS0091P	925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #	
	BACR19D16 of RpCl-98 library from Drosophila melanogaster (fruit	
	fly), genomic survey sequence.	

XX (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (INSR) INST ROUSSY GUSTAVE.

XX Mastangeli R, Romagnani S, Triebel F;
 DR MPI, 1999-080953/07.
 XX P-PSDB; AAM88361.

PT New lymphocyte activation gene 3 (LAG-3) splice variants - which
 can be used as immunomodulators

XX Claim 1; Page 26-27; 49pp; English.

CC This cDNA encodes LAG-3V3 (see AAM88361), a splice variant of human
 CC lymphocyte activation gene 3 (LAG-3) protein. LAG-3V3 derives from
 CC cleavage of the nuclear transcript at a different polyA site
 CC located approximately 170 bp downstream to the 5' end of intron 5.
 CC The retained intron 5 sequence contains an in-frame stop codon. The
 CC resulting mRNA encodes a truncated soluble protein which contains
 CC D1, D2, D3 and 8 new amino acid residues. LAG-3V2 cDNA was
 CC obtained by RT-PCR amplification (see AAX06759 and AAX06762) of
 CC peripheral blood mononuclear cell RNA. The invention provides 3
 CC splice variants (see AAM88359-61) of LAG-3 and nucleotide sequences
 CC encoding them (see AAX06754-56). The LAG-3 variants can be used in,
 CC or for the manufacture of, therapeutic compositions used to treat
 CC immune-related pathologies (claimed), in particular T11-dependent
 CC diseases such as Hashimoto's thyroiditis, type I diabetes mellitus,
 CC multiple sclerosis, Crohn's disease, rheumatoid arthritis, acute
 CC allograft rejection, acute graft-versus-host disease, Grave's
 CC ophthalmopathy, cerebral malaria, Lyme arthritis, reactive arthritis,
 CC HCV-induced chronic hepatitis, primary sclerosing colangitis, contact
 CC dermatitis, unexplained recurrent abortion, aplastic anaemia and
 CC Helicobacter pylori induced gastric arthritis. LAG-3 variants can
 CC also be used for the manufacture of immunomodulator compounds
 CC (claimed), which can mimic or alter the biological function of
 CC LAG-3.

XX Sequence 1468 BP; 231 A; 550 C; 404 G; 283 T; 0 other;

XX Query Match 100.0%; Score 500; DB 20; Length 1468;

XX Best Local Similarity 100.0%; Pred. No. 1,9e-105; Indels 0; Gaps 0;

XX Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGGCTGCTGATGTCAGGCTTTCAGGCTTCTGATGTCGCTGATGTC 60
 DB 1 TCAGGCTGCTGATGTCAGGCTTTCAGGCTTCTGATGTCGCTGATGTC 60
 QY 61 CTTGCCACCTCTCTCTCAAGGCTTCTCTGTCCTCTCTCTCTCTCTCTCT 120
 DB 61 CTTGCCACCTCTCTCTCAAGGCTTCTCTGTCCTCTCTCTCTCTCTCTCT 120
 QY 121 CACCT 180
 DB 121 CACCT 180
 QY 181 TTTTCTGACCT 240
 DB 181 TTTTCTGACCT 240
 QY 241 CTTGAGTCT 300
 DB 241 CTTGAGTCT 300
 QY 301 AGCCAGGCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
 DB 301 AGCCAGGCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
 QY 361 GCAGCCCAATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
 DB 361 GCAGCCCAATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420

OY 421 AGCATCAGCCAGACAGTGGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 480
 DB 421 AGCATCAGCCAGACAGTGGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 480
 QY 481 CTCACCCGCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 500
 DB 481 CTCACCCGCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 500

RESULT 2
 ID AAX06755 standard; cDNA; 1629 BP.
 XX AAX06755
 XX AAX06755;
 XX 26-APR-1999 (first entry)

Human lymphocyte activation gene 3 splice variant LAG-3V2 cDNA.
 DE Lymphocyte activation gene 3; LAG-3; LAG-3V2; splice variant;
 XX human; immunomodulator; Hashimoto's thyroiditis;
 KW type I diabetes mellitus; multiple sclerosis; Crohn's disease;
 KW rheumatoid arthritis; allograft rejection;
 KW graft-versus-host disease; Grave's ophthalmopathy; abortion;
 KW cerebral malaria; Lyme arthritis; reactive arthritis; hepatitis;
 KW primary sclerosing colangitis; dermatitis; aplastic anaemia;
 KW gastric arthritis; therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 297..1565
 FT /tag= a

XX W09858059-A1.

XX 23-DEC-1998.

XX 03-JUN-1998; 98WO-EP03307.

XX 18-JUN-1997; 97EP-0401404.

XX (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (INSR) INST ROUSSY GUSTAVE.

XX Mastangeli R, Romagnani S, Triebel F;
 DR MPI, 1999-080953/07.

XX P-PSDB; AAM88360.

XX New lymphocyte activation gene 3 (LAG-3) splice variants - which
 can be used as immunomodulators

XX Claim 1; Page 24-25; 49pp; English.

CC This cDNA encodes LAG-3V2 (see AAM88360), a splice variant of human
 CC lymphocyte activation gene 3 (LAG-3) protein. LAG-3V2 lacks exon
 CC 6 due to the ligation of the donor site of intron 5 to the acceptor
 CC site of intron 6. There is no frame shift. The LAG-3V2 protein is
 CC a transmembrane protein which does not contain the D4 domain.
 CC LAG-3V2 cDNA was obtained by RT-PCR amplification (see AAX06759 and
 CC AAX06762) of peripheral blood mononuclear cell RNA. The invention
 CC provides 3 splice variants (see AAM88359-61) of LAG-3 and nucleotide
 CC sequences encoding them (see AAX06754-56). The LAG-3 variants can
 CC be used in, or for the manufacture of, compositions used to treat
 CC immune-related pathologies (claimed), in particular T11-dependent
 CC diseases such as Hashimoto's thyroiditis, type I diabetes mellitus,
 CC multiple sclerosis, Crohn's disease, rheumatoid arthritis, acute
 CC allograft rejection, acute graft-versus-host disease, Grave's
 CC ophthalmopathy, cerebral malaria, Lyme arthritis, reactive arthritis,
 CC HCV-induced chronic hepatitis, primary sclerosing colangitis, contact
 CC dermatitis, unexplained recurrent abortion, aplastic anaemia and

DT 24-SEP-2001 (first entry)

Human small cell lung cancer associated gene, LAG-3.

KM Human; small cell lung cancer; therapy; hCAP; nucleic acid; NA;
KM melanoma; cancer; colon; breast; head; neck; transitional cancer;
KM leiomyosarcoma; lymphocyte activation gene-3; LAG-3; synovial sarcoma;
KM cytosolic; ds.

OS Homo sapiens.

PN WO200153349-A2.

PD 26-JUL-2001.

PF 19-JAN-2001; 2001WO-US02015.

PR 21-JAN-2000; 2000US-0489101.

PA (LUDW-) LUDWIG INST CANCER RES.

PA (CORR) CORNELL RES FOUND INC.

PI Stockert E, Scanlan MJ, Jager

DR WPI; 2001-457597/49.

PT Isolated polypeptide, used to treat or prophage a disorder
PT Characterized by expression of a hCAP e.g. cancer, is encoded by an
PT Isolated nucleic acid comprising an NA Group 3 or 4 molecule -
XX
XS Claim 57, Page 108-109, 152pp; English.

CC The invention relates to nucleic acids and encoded polypeptides which
CC are cancer associated antigens expressed in patients afflicted with
CC small cell lung cancer. The molecules provided by the invention can be
CC used in the diagnosis, monitoring, research or treatment of conditions
CC characterised by the expression of one or more cancer associated
CC antigens. The polypeptide is used to treat a disorder characterised by
CC expression of a hCAAP, and determine regression, progression or onset
CC of a condition characterised by expression of an abnormal amount of a
CC protein encoded by a nucleic acid (NA) group 1 molecule. The disorders
CC are small and non-small cell lung cancer, melanoma, colon, breast, head
CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
CC The present sequence is a DNA encoding human lymphocyte activation
CC gene-3 (LAG-3). This small cell lung cancer associated gene is
CC designated as NY-SCLC-12.

Sequence 1872 BP; 300 A; 673 C; 523 G; 376 T; 0 other;

```
Query match      100.0%;  Score 500;  DB 22;  Length 1872;
```

```
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	TTGGGCGTCGTCATATGCGCCAGACCTTTCACGCTTCCGTCGGATATCCGGGCTCGTGGTATC	60
Db	1	TGAGGTGCTCATATGCTCCAGCTTTCACGCTTTCCTCGATATCCGGGCTCTGGTATATC	60
QY	61	CGTCCCGACACCTCTCTCCCAAGGCCCTCTCGTGGTCTCCCTCTCTCTAGAACCCCTTCCTC	120
Db	61	CGTCCCGACACCTCTCTCCCAAGGCCCTCTCGTGGTCTCCCTCTCTCTAGAACCCCTTCCTC	120
QY	121	CACCTCCCTCTCTGACAGACTCTCTCTTACCCGCCACCCGCCACCCACCTGCCCCCTTTC	180
Db	121	CACCTCCCTCTCTGAGAACTCTCTCTTACCCCCACCCGCCACCCACCTGCCCCCTTTC	180
QY	181	TTTTCTGACCTCTCTTTTGAGAGGCTCGACGCTGCGCCAGACCATAGAGAGATGTGGAGG	240
Db	181	TTTTCTGACCTCTCTTTTGAGAGGCTCGACGCTGCGCCAGACCATAGAGAGATGTGGAGG	240
QY	241	CTCAGTTCCCTGGGCTTGCCTGTTCTTCGAGAGCGCTTTTGAGTGAGCTCCAGTGAAGCCTTCAC	300
Db	241	CTCAGTTCCCTGGGCTTGCCTGTTCTTCGAGAGCGCTTTTGAGTGAGCTCCAGTGAAGCCTTCAC	300

QY	361	AGCAGGCGGCTAGGTCCTCCGGGTGAGTGGGCGCCAGAGAGGGGCTCTGACCCAGCTCCCT	360
QY	301	AGCCAGGGCTAGGTCCTCCGGGTGAGTGGGCGCCAGAGAGGGGCTCTGACCCAGCTCCCT	300
Db	301	AGCCAGGGCTAGGTCCTCCGGGTGAGTGGGCGCCAGAGAGGGGCTCTGACCCAGCTCCCT	360
QY	361	GCAGCCCCCAATCCCTCCAGAGATCTGCAGCTCTCTGGAGAGCAGAGGGTCACCTTGGC	420
Db	361	GCAGCCCCCAATCCCTCCAGAGATCTGCAGCTCTCTGGAGAGAGAGAGGGTCACCTTGGC	420
QY	421	AGCATATAGCCAGACAGTGGGCCGCTCCGCTGCGCGCCCGGCATATCCCTGGGCCCCGGCC	480
Db	421	AGCATATAGCCAGACAGTGGGCCGCTCCGCTCCGCGCCCGGCATATCCCTGGGCCCCGGCC	480
QY	481	CTTCACCCGGGCGGCGCCCTCC	500
Db	481	CTTCACCCGGGCGGCGCCCTCC	500

RESULT 5

ID AAX06754 standard; cDNA; 2279 BP.

AC AAX06754

26-APR-1999 (first entry)

Human lymphocyte activation gene 3 splice variant LAG-3V1 cDNA.

AA
KW
Lymphocyte activation gene 3: LAG-3: LAG-3V1: splice variant

type I diabetes mellitus: multiple sclerosis: Crohn's disease:
KW human; influenza; hashimoto's thyroiditis;

KW graft-versus-host disease: Grave's ophthalmopathy: abortion:
 KW live-and-let-be, religious rejection;
 KW

primary sclerosing cholangitis; dermatitis: aplastic anaemia: kw cerebral malaria; Lyme arthritis; reactive arthritis; hepatitis; kw

gastric antritis; therapy; ss.

US Homo sapiens
XY

EH	Key	Location/Qualifiers
ET	CNS	397 1MD

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

PN W09858059-A1
XY

PD 23-DEC-1998.

PF 03-JUN-1998; 98WO-EP033307.

PR 18-JUN-1997; 97EP-0401404.

PA (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.

PA (INSR) INST ROUSSY GUSTAVE.

PI Mastrangeli R, Romagnani S, Triebel F,

DR WPI; 1999-080953/

XX

PT can be used as immunomodulators

XX This cDNA encodes LAG-3V1 (see AAW883159), a splice variant of human
PS lymphocyte activation gene 3 (LAG-3) protein. LAG-3V1 is derived
CC from the retention of intron 4, i.e. cleavage at the donor and
CC acceptor sites flanking intron does not occur. An in-frame stop
CC codon located after 8 codons in the retained intron 4 leads to a 36
CC kDa truncated soluble LAG-3V1 protein containing D1, D2 and 8 new
CC amino acid residues. LAG-3V1 cDNA was obtained by RT-PCR
CC amplification (see AAX06757-58) of peripheral blood mononuclear cells.

CC RNA. The invention provides 3 splice variants (see AAM8359-61) of
 CC LAG-3 and nucleotide sequences encoding them (see AAX06754-56). The
 CC LAG-3 variants can be used in, or for the manufacture of,
 CC pharmaceutical or therapeutic compositions which are used to treat
 CC immune-related pathologies (claimed), in particular Th1-dependent
 CC diseases such as Hashimoto's thyroiditis, type 1 diabetes mellitus,
 CC multiple sclerosis, Crohn's disease, rheumatoid arthritis, acute
 CC allograft rejection, acute graft-versus-host disease, Grave's
 CC ophthalmopathy, cerebral malaria, Lyme arthritis, reactive arthritis,
 CC HCV-induced chronic hepatitis, primary sclerosing colangitis, contact
 CC dermatitis, unexplained recurrent abortion, aplastic anaemia and
 CC Helicobacter pylori-induced gastric antritis. LAG-3 variants can
 CC also be used for the manufacture of immunomodulator compounds
 CC (claimed), which can mimic or alter the biological function of
 CC LAG-3.

XX Sequence 2279 BP; 360 A; 832 C; 589 G; 498 T; 0 other;

Query Match 100.0%; Score 500; DB 20; Length 2279;
 Best Local Similarity 100.0%; Pred. No. 2.1e-105;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCAGGTCCTCCGATCTGCGCCAGCTTTCAGATCTCCCTGATGCCGCTGCTGATC 60
 DB 1 TCAGGTCCTCCGATCTGCGCCAGCTTTCAGATCTCCCTGATGCCGCTGCTGATC 60
 OY 61 CTTCCCAACCTCTCTCCAGAGCCCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTC 120
 DB 61 CTTCCCAACCTCTCTCCAGAGCCCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTC 120
 OY 121 CACCTCCCTCTCTGAGAACTTCTCTTACCCCAACCCCAACCTCTCTCTCTCTC 180
 DB 121 CACCTCCCTCTCTGAGAACTTCTCTTACCCCAACCCCAACCTCTCTCTCTCTC 180
 OY 181 TTTTCTGACCTCTTTTGAGAGGCTCAGCGCTGCCAGACATAGAGAGATGSGAGG 240
 DB 181 TTTTCTGACCTCTTTTGAGAGGCTCAGCGCTGCCAGACATAGAGAGATGSGAGG 240
 OY 241 CTCAGTTCTGAGGCTGCTGTTTCTGAGAGCCCTTTGAGGCTCTCCAGTGAAGCTCTCC 300
 DB 241 CTCAGTTCTGAGGCTGCTGTTTCTGAGAGCCCTTTGAGGCTCTCCAGTGAAGCTCTCC 300
 OY 301 AGCCAGGGGCTGAGTCTCCGCTGCTGAGGCTCTGAGGCTCTGAGTCTCTCCCT 360
 DB 301 AGCCAGGGGCTGAGTCTCCGCTGCTGAGGCTCTGAGGCTCTGAGTCTCTCCCT 360
 OY 361 GCAGCCCAACATCCCTCCAGATCTCAGCTTCTGCAAGAGAGAGGCTCACTTGGC 420
 DB 361 GCAGCCCAACATCCCTCCAGATCTCAGCTTCTGCAAGAGAGAGGCTCACTTGGC 420
 OY 421 AGCATCAGCAGAGAGTGGCCGCTGCTGAGGCTCTGAGGCTCTGAGGCTCTGAGG 480
 DB 421 AGCATCAGCAGAGAGTGGCCGCTGCTGAGGCTCTGAGGCTCTGAGGCTCTGAGG 480
 OY 481 CTCACCGGCGGCGGCTCTCC 500
 DB 481 CTCACCGGCGGCGGCTCTCC 500

RESULT 6
 AAQ12871 standard; DNA; 1164 BP.

XX AAQ12871;

XX 15-OCT-1991 (first entry)

DE Polyhedrin-soluble LAG-3 fusion construct.

XX human cytotoxic T-lymphocyte; LAG-3S; HIV; CD4; AIDS; ss.

XX Homo sapiens.

FH Key Location/Qualifiers
 FT sig_peptide 22..105
 FT mat_peptide 106..1161
 FT /tag= a
 FT /tag= b
 FT /product= LAG-3S fused to polyhedrin fragment

XX W09110682-A.

XX 25-JUL-1991.

XX 08-JAN-1991; 91WO-FR00009.

XX 08-JAN-1990; 90FR-0000126.

XX (INRA) INSERM INST NAT SANTE.

XX (INSR) INST ROUSSY G.

XX Herceud T, Triebel F;

XX WPI: 1991-237986/32.

XX P-PSDB; AAR13272.

XX Human lymphocyte membrane protein and DNA encoding it (LAG-3) -
 XX useful for antibodies and pharmaceutical products used to treat
 XX HIV-infected humans

XX Example 11; Page 49-50; 69pp; French.

XX FDC cDNA was isolated from a natural cytotoxic lymphocyte cDNA
 XX library constructed in lambda gt10. It encodes LAG-3, a
 XX membrane-bound protein. A soluble form (LAG-3S) of the protein can
 XX be produced in insect cells (i.e. Spodoptera frugiperda Sf9 cells)
 XX by removing the transmembrane region-coding sequence. This
 XX construct encodes the first 3 extracellular 19-11 kDa domains of
 XX LAG-3 (V domain and two C2 domains) fused to a 17 amino acid
 XX C-fragment of the insect protein polyhedrin.
 XX See also AAQ12869-70.

XX Sequence 1164 BP; 182 A; 426 C; 340 G; 216 T; 0 other;

Query Match 58.2%; Score 291; DB 12; Length 1164;
 Best Local Similarity 100.0%; Pred. No. 1.4e-57;
 Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 210 GCTGCCAGACCATAGAGAGATGAGAGGCTCAGTCTGCTGCTCTGTTCTGAG 269
 DB 1 GCTGCCAGACCATAGAGAGATGAGAGGCTCAGTCTGCTGCTCTGTTCTGAG 60
 OY 270 CGCTTTGGGTGCTCAGTGAAGCTCTCCAGCCAGGGGCTGAGGCTCCGGTGTGG 329
 DB 270 CGCTTTGGGTGCTCAGTGAAGCTCTCCAGCCAGGGGCTGAGGCTCCGGTGTGG 120
 OY 330 GCCCAGAGGGGCTCTCCAGCTCCCTCCAGCCCAACATCCCTCCAGATCTC 389
 DB 330 GCCCAGAGGGGCTCTCCAGCTCCCTCCAGCCCAACATCCCTCCAGATCTC 180
 OY 390 AGCTTCTGGAAGAGAGAGGCTCTGAGAGTCAAGTCAAGAGAGAGGCTCCGCT 449
 DB 181 AGCTTCTGGAAGAGAGAGGCTCTGAGAGTCAAGTCAAGAGAGAGGCTCCGCT 240
 OY 450 GCCGCCCGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
 DB 241 GCCGCCCGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 291

RESULT 7

XX AAQ12870 standard; DNA; 999 BP.

XX AAQ12870;

XX 15-OCT-1991 (first entry)

```

XX Promoter sequence for LAG-3 expression.
DE human cytotoxic T-lymphocyte; HIV; CD4; AIDS; ss.
XX Homo sapiens.
XX MO9110682-A.
XX 25-JUL-1991.
XX 08-JAN-1991; 91WO-FR00009.
XX 08-JAN-1990; 90FR-0000126.
XX (INRM ) INSERM INST NAT SANTE.
XX (INSR ) INST ROUSSY G.
XX Herceud T, Triebel F;
XX WPI: 1991-237986/32.
XX Human lymphocyte membrane protein and DNA encoding it (LAG-3) -
XX useful for antibodies and pharmaceutical products used to treat
XX HIV-infected humans
XX Claim 26; Page 48; 69pp; French.
XX This sequence is a suitable promoter for expression of LAG-3. It is
XX also claimed as a DNA sequence which encodes a polypeptide fragment
XX of LAG-3 although no correspondence amino acid sequence is given.
XX See also AAQ12869 and AAQ12871.
XX Sequence 999 BP; 214 A; 278 C; 302 G; 205 T; 0 other;
SQ
Query Match 44.2%; Score 221; DB 12; Length 999;
Best Local Similarity 99.6%; Pred. No. 1.5e-41;
Matches 232; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 TCAGGCTGCTGATCTGCGCCAGCTTTCAGCTTCTCTGATTCGCGCTGTCATC 60
DB 768 TCAGGCTGCTGATCTGCGCCAGCTTTCAGCTTCTCTGATTCGCGCTGTCATC 827
QY 61 CCTCCGACCTCTCTCCAAAGGCGCTCCGCTGCTCCCTCTCTGAGACCCCTTCTC 120
DB 828 CCTCCGACCTCTCTCCAAAGGCGCTCCGCTGCTCCCTCTCTGAGACCCCTTCTC 887
QY 121 CACCTCCCTCTCTGAGAACTTCTCTTACCCGCCACCCGCCACCACTGCCCTTTC 180
DB 888 CACCTCCCTCTCTGAGAACTTCTCTTACCCGCCACCCGCCACCACTGCCCTTTC 946
QY 181 TTTTCTGACCTCTTTTGGAGGCTCAGCGCTGCCAGACATGAGAGATG 233
DB 947 TTTTCTGACCTCTTTTGGAGGCTCAGCGCTGCCAGACATGAGAGATG 999
RESULT 8
AAV07462
ID AAV07462 standard; DNA: 447 BP.
XX AAV07462;
XX 01-OCT-1998 (first entry)
XX DNA encoding the first domain, D1, of the LAG-3 protein.
XX Domain D1; lymphocyte activating gene 3; LAG-3; reduced affinity;
XX class II molecule; immunomodulator; treatment; autoimmune disease;
XX graft rejection; anticancer immunotherapy; ss.
XX Homo sapiens.
XX MO9823741-A1.

```

```

XX 04-JUN-1998.
XX 25-NOV-1997; 97WO-FR02126.
XX 28-NOV-1996; 96FR-0014608.
XX (ISFE ) ARS APPLIED RES SYSTEMS HOLDING NV.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (INSR ) INST ROUSSY GUSTAVE.
XX El Tayar N, Huard B, Mastrangeli R, Triebel F;
XX WPI: 1998-322721/28.
XX P-PSDB; AAW51846.
XX New mutants of lymphocyte activating gene-3 protein - comprise
XX extracellular domains, having reduced interaction with Class II
XX molecules, useful as immuno-modulator for treating, e.g. auto-immune
XX disease
XX Disclosure; Fig 2; 62pp; French.
XX The present sequence encodes the first extracellular domain, D1, of a
XX protein encoded by lymphocyte activating gene 3 (LAG-3). The sequence is
XX used to create mutants AAW62901-02. The mutants have reduced, or no,
XX affinity for Class II molecules. These mutants are useful as
XX immunomodulators for treating autoimmune disease, for controlling graft
XX rejection and in anticancer immunotherapy (e.g. of leukaemia and
XX melanoma), and also to produce antagonists or agonists of the
XX interaction between LAG-3 and Class II major histocompatibility
XX molecules. Cells in which the mutants are exposed at the surface are
XX useful in binding/adhesion assays to study binding to particular
XX ligands.
SQ
Sequence 447 BP; 49 A; 188 C; 152 G; 58 T; 0 other;
Query Match 40.8%; Score 204; DB 19; Length 447;
Best Local Similarity 100.0%; Pred. No. 9.9e-38;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 297 CTCGAGCCAGGAGGCGGAGGTCGGTGTGGGCGCAGAGAGGAGGCTCTGCGCAGCTC 356
DB 1 CTCGAGCCAGGAGGCGGAGGTCGGTGTGGGCGCAGAGAGGAGGCTCTGCGCAGCTC 60
QY 357 CCTGCAAGCCCAACATCCCTCTCAGAGATCTCAGCTTCTGCGAAGAGCAGGGGTCTACT 416
DB 61 CCTGCAAGCCCAACATCCCTCTCAGAGATCTCAGCTTCTGCGAAGAGCAGGGGTCTACT 476
QY 417 TGGCAGCATCAGCAGACAGTGGCGCGCGCTGCGCCGCCCATCCCTGCGCCGCC 476
DB 121 TGGCAGCATCAGCAGACAGTGGCGCGCGCTGCGCCGCCCATCCCTGCGCCGCC 180
QY 477 GGCCCTCACCAGGCGGCGCCCTCC 500
DB 181 GGCCCTCACCAGGCGGCGCCCTCC 204
RESULT 9
ABL33153/C
ID ABL33153 standard; DNA: 5689 BP.
XX ABL33153;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1126.
XX Human: immune system disease; cytosine methylation; antileukemic;
XX antileukosclerotic; antianemic; cytoskeletal; neoplastic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antineumatic; antidiabetic; antidiabetic; antipsoriatic;
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

```

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 OS Homo sapiens.
 PN W0200200928-A2.
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS Claim 1; SEQ ID NO 1126; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SQ Sequence 5689 BP; 1238 A; 120 C; 1590 G; 2741 T; 0 other;
 XX
 Query Match 40.8%; Score 204; DB 24; Length 5689;
 Best Local Similarity 81.2%; Pred. No. 1.5e-37;
 Matches 237; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 YY 1 TCAGGCTGCTGATCTGCCAGCTTTCAGCTTCTCTGATTCGGGCTTGGTCATC 60
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 570 TCAACTACCTATCTATCCCACTTTCACCTTCTCTAAATTCGACCTATATCATC 511
 YY 61 CCTGCCACCTCTCTCCAGGCGCTCTCTGTCCTCTCTCTAGAACCTTCTCTC 120
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 510 CCGCCACCTCTCTCCAAACCGCTCTCTATCTCTCTTAAAGCCCTTCTC 451
 YY 121 CACCTCCCTCTGCGAGAACTCTCTTACCCGCCAGCCGCCAGCAGCTTCTC 180
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 450 CACCTCCCTCTCTCAAAACCTCTCTTACCCGCCAGCCGCCAGCAGCTTCTC 391
 YY 181 TTTTCTGACCTCTCTTTGGAGGCTTCAGGCGTCCAGACCATAGAGAGATGGGAGG 240
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 390 TTTTCTAACTCTCTTTTAAAGCATCAAGCTAACCAAAATATAAATAAATAA 331
 YY 241 CTGAGTCTCTGGGCTCTCTCTCTGAGCGCTTGGTGGCTCTCAAGTAA 292
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 330 CTCAATTCCTAAACTTACTTCTACAAACCGCTTAAATTAATCACTCAATA 279
 XX
 RESULT 10
 ID ABL34551/c
 AC ABL34551 standard; DNA; 5689 BP.
 XX
 AC ABL34551;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human metastasis associated gene SEQ ID NO: 104.
 XX

KW Metastasis associated gene; cytostatic; gene therapy; cancer;
 KW cytosine methylation; gene; ds.
 OS Homo sapiens.
 PN W0200177376-A2.
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03970.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-010922/01.
 XX
 PT New nucleic acid derived from chemically treated metastasis genes,
 PT useful for diagnosis of cancers by analysis of cytosine methylation,
 PT also for treatment -
 PS Claim 1; SEQ ID NO 104; 23pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of human metastasis associated
 CC genes which are modified by cytosine methylation. The sequences can be
 CC used in the diagnosis and treatment of cancer. The present sequence is
 CC one of the genes of the invention.
 CC
 SQ Sequence 5689 BP; 1238 A; 120 C; 1590 G; 2741 T; 0 other;
 XX
 Query Match 40.8%; Score 204; DB 24; Length 5689;
 Best Local Similarity 81.2%; Pred. No. 1.5e-37;
 Matches 237; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 YY 1 TCAGGCTGCTGATCTGCCAGCTTTCAGCTTCTCTGATTCGGGCTTGGTCATC 60
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 570 TCAACTACCTATCTATCCCACTTTCACCTTCTCTAAATTCGACCTATATCATC 511
 YY 61 CCTGCCACCTCTCTCCAGGCGCTCTCTGTCCTCTCTCTAGAACCTTCTCTC 120
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 510 CCGCCACCTCTCTCCAAACCGCTCTCTATCTCTCTTAAAGCCCTTCTC 451
 YY 121 CACCTCCCTCTGCGAGAACTCTCTTACCCGCCAGCCGCCAGCAGCTTCTC 180
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 450 CACCTCCCTCTCTCAAAACCTCTCTTACCCGCCAGCCGCCAGCAGCTTCTC 391
 YY 181 TTTTCTGACCTCTCTTTGGAGGCTTCAGGCGTCCAGACCATAGAGAGATGGGAGG 240
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 390 TTTTCTAACTCTCTTTTAAAGCATCAAGCTAACCAAAATATAAATAAATAA 331
 YY 241 CTGAGTCTCTGGGCTCTCTCTCTGAGCGCTTGGTGGCTCTCAAGTAA 292
 YY 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 330 CTCAATTCCTAAACTTACTTCTACAAACCGCTTAAATTAATCACTCAATA 279
 XX
 RESULT 11
 ID ABL33152
 AC ABL33152 standard; DNA; 5689 BP.
 XX
 AC ABL33152;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1125.
 XX
 DE Human immune system associated gene; cytosine methylation; antiasthmatic;
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation

PS Claim 1; SEQ ID NO 1125; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

CC Sequence 5689 BP; 1151 A; 120 C; 1710 G; 2708 T; 0 other;

Query Match 21.3%; Score 106.4; DB 24; Length 5689;
 Best Local Similarity 60.3%; Pred. No. 3.5e-15;
 Matches 176; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1 TCAGGCTGCGTGCATGCGCCAGCTTCCAGCTTCCCTGATTCGGCCCTGCTGATC 60

DB 5120 TTAGGTTGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTATP 5179

QY 61 CTTCCCAACCTCTCTCCAGAGCCCTCTCCGTCCTCTCTCTCTCTCTCTCTCTCTCTC 120

DB 5180 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5239

QY 121 CACCTCCCTCTCTGCAAGACTTCTCTTACCCCAACCCCAACCCCAACCCCTTCTC 180

DB 5240 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5299

QY 181 TTTTCTGACCTCTTTGAGAGGCTCAGAGCTGCCAGACATGAGAGATGAGAGG 240

DB 5300 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5359

QY 241 CTCAGTTCTCTGAGGCTGCTTTCTGACAGCCGCTTGGAGGCTCCAGTGA 292

DB 5360 TTTAGTTTGGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTAA 5411

RESULT 12

ID ABL34550 standard; DNA; 5689 BP.

AC ABL34550;

DT 26-MAR-2002 (first entry)

DE Human metastasis associated gene SEQ ID NO: 103.

KW Metastasis associated gene; cytostatic; gene therapy; cancer;
 KW cytosine methylation; gene; ds.

OS Homo sapiens.

PN WO20017376-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP03970.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-010922/01.

PT New nucleic acid derived from chemically treated metastasis genes,
 PT useful for diagnosis of cancers by analysis of cytosine methylation,
 PT also for treatment

PS Claim 1; SEQ ID NO 103; 23pp + Sequence Listing; English.

CC The present invention provides a number of human metastasis associated
 CC genes which are modified by cytosine methylation. The sequences can be
 CC used in the diagnosis and treatment of cancer. The present sequence is
 CC one of the genes of the invention.

CC Sequence 5689 BP; 1151 A; 120 C; 1710 G; 2708 T; 0 other;

Query Match 21.3%; Score 106.4; DB 24; Length 5689;
 Best Local Similarity 60.3%; Pred. No. 3.5e-15;
 Matches 176; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1 TCAGGCTGCGTGCATGCGCCAGCTTCCAGCTTCCCTGATTCGGCCCTGCTGATC 60

DB 5120 TTAGGTTGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTATP 5179

QY 61 CTTCCCAACCTCTCTCCAGAGCCCTCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCTC 120

DB 5180 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5239

QY 121 CACCTCCCTCTCTGCAAGACTTCTCTTACCCCAACCCCAACCCCAACCCCTTCTC 180

DB 5240 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5299

QY 181 TTTTCTGACCTCTTTGAGAGGCTCAGAGCTGCCAGACATGAGAGATGAGAGG 240

DB 5300 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5359

QY 241 CTCAGTTCTCTGAGGCTGCTTTCTGACAGCCGCTTGGAGGCTCCAGTGA 292

DB 5360 TTTAGTTTGGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTAA 5411

RESULT 13

ID AAS85793/C standard; CDNA; 3267 BP.

AC AAS85793;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #21597.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;


```
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB: ABG21606.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID NO 21597; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3267 BP; 572 A; 914 C; 970 G; 811 T; 0 other;
XX
XX Query Match 13.4%; Score 67.2; DB 23; Length 3267;
XX Best Local Similarity 95.8%; Pred. No. 3e-06; 3; Indels 0; Gaps 0;
XX Matches 69; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 429 CCAACAGTGGCGCGCGCGCGCGCGCGCATCCCTGCGCCCGGCGCTACCCG 488
XX |||
XX DB 2250 CGCGCCAGTGGCGCGCGCGCGCGCGCGCGCATCCCTGCGCCCGGCGCTACCCG 2191
XX |||
XX QY 489 GCGGCGCCCTCC 500
XX |||
XX DB 2190 GCGGCGCCCTCC 2179
XX |||
XX
XX RESULT 14
XX AAS86048
XX ID AAS86048 standard; cDNA; 5912 BP.
XX AC AAS86048;
XX AC AAS86048;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #21852.
```

```
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB: ABG21861.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID NO 21852; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5912 BP; 1540 A; 1587 C; 1551 G; 1234 T; 0 other;
XX
XX Query Match 13.4%; Score 67.2; DB 23; Length 5912;
XX Best Local Similarity 95.8%; Pred. No. 3.3e-06; 3; Indels 0; Gaps 0;
XX Matches 69; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 429 CCAACAGTGGCGCGCGCGCGCGCGCGCGCATCCCTGCGCCCGGCGCTACCCG 488
XX |||
XX DB 1018 CGCGCCAGTGGCGCGCGCGCGCGCGCGCGCATCCCTGCGCCCGGCGCTACCCG 1077
XX |||
XX QY 489 GCGGCGCCCTCC 500
XX |||
XX DB 1078 GCGGCGCCCTCC 1089
XX |||
XX
XX RESULT 15
XX AAZ17263
XX ID AAZ17263 standard; cDNA; 1337 BP.
XX AC AAZ17263;
XX AC AAZ17263;
XX DT 12-OCT-1999 (first entry)
```

XX Human gene expression product cDNA sequence SEQ ID NO:4735.

XX Human; gene: gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

XX MO9938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99MO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

XX Lamson G, Leshkowitz D, Pol D, Randazzo F, Reinhard C;

XX Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are

XX differentially expressed in different cell types

XX Claim 1: Page 2250-2251; 2479pp; English.

XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA12532 to AA17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA12532 to AA17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensic, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;

XX Query Match 10.0%; Score 49.8; DB 20; Length 1337;

XX Best Local Similarity 26.5%; Pred. No. 0.025;

XX Matches 132; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

XX 2 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 61

XX 592 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 61

XX 62 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 121

XX 652 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 711

QY 122 ACCTCCCTCTGAGAACTTCTCCTTACCCTCCACACCCACCACTGCTTCT 181
 Db 712 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 771
 QY 182 TTTCTGACCTCTTTTGGAGGCTCAGGCTGCCAGACCATAGAGATGTGGAGGC 241
 Db 772 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 831
 QY 242 TCAGTTCCTGAGCTTGTCTGTTTCTGACGCTTTGGGTGCTCAAGCTTCTCA 301
 Db 832 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 891
 QY 302 GCGAGGGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
 Db 892 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 951
 QY 362 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 421
 Db 952 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 1011
 QY 422 GATCAGCCAGACAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 481
 Db 1012 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 1071
 QY 482 TCAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 500
 Db 1072 CAGGCTGCTGATGTGCGCAGCTTTCAGCTTTCGATTCGGGCTTGATCC 1090

Search completed: October 10, 2002, 15:17:48
 Job time: 102 secs


```
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreyer, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; Query Match 12.0%; Score 60; DB 1; Length 10596;
; Best Local Similarity 50.6%; Pred. No. 0.00037;
; Matches 170; Conservative 0; Mismatches 165; Indels 1; Gaps 1;
;
; QY 131 GCCAGATAGAGCGCCGCGGCGGCGGAGTCCGCGCGCTGCGCGCGCGCG 190
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2601 GCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2542
; QY 191 AGGACCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 250
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2541 GCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2482
; QY 251 CCCGAGATAGAGCGCCGCGGCGGCGCGCGCGCGCGCGCGCGCGCG 310
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2481 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2422
; QY 311 CGCGATGTCG-ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2421 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2362
; QY 370 CGAGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2361 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2302
; QY 430 ACACGAGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2301 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2266
```

```
RESULT 7
US-08-087-783A-15/c
; Sequence 15, Application US/08087783A
```

```
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
;
; US-08-087-783A-15
;
; Query Match 12.0%; Score 60; DB 1; Length 10596;
; Best Local Similarity 50.6%; Pred. No. 0.00037;
; Matches 170; Conservative 0; Mismatches 165; Indels 1; Gaps 1;
;
; QY 131 GCCAGATAGAGCGCCGCGGCGGCGGAGTCCGCGCGCTGCGCGCGCG 190
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2601 GCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2542
; QY 191 AGGACCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 250
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2541 GCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2482
; QY 251 CCCGAGATAGAGCGCCGCGGCGGCGCGCGCGCGCGCGCGCGCGCG 310
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2481 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2422
; QY 311 CGCGATGTCG-ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2421 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2362
; QY 370 CGAGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2361 CTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2302
; QY 430 ACACGAGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465
; | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2301 CTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2266
```

```
RESULT 8
```


GENERAL INFORMATION:

APPLICANT: Genentech, Inc., Godowski, Paul J., Loket, Natalie A., Mark, Melanie H
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648
FILING DATE: 19930517
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885971
FILING DATE: 18-MAY-92
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755,779P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US93-04648-15

Query Match 12.0%; Score 60; DB 5; Length 10596;

Best Local Similarity 50.6%; Pred. No. 0.00037;

Matches 170; Conservative 0; Mismatches 165; Indels 1; Gaps 1;

QY 131 GCCCAGATAGAGCCCGGGGGGCGGCTGAGAGTCCCGCGGCTGCTGCGCGCG 190
DB 2601 GCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2542
QY 191 AGAGCCGACCCGCGCTGCGCGCGATGCTGAGTGGGGCGCGCCATGAGAGGATTA 250
DB 2541 GCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2482
QY 251 CCCGAGATAGAGCCCGCGGGGCGAGCTGCTGACAGCGCGCGCGCTGCGAG 310
DB 2481 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2422
QY 311 CGCATGCTGC-ACGCGGCTACTGGAACCTTCGCTGCGCGCATACCTGCTCT 369
DB 2421 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2362
QY 370 CCAAGCAGAGCGCGCGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 429
DB 2361 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2302
QY 430 ACACGATATGCGCGCGTGGGACCGCGCTCCAGCT 465
DB 2301 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2266

RESULT 11
US-09-128-155-16

Sequence 16, Application US/09128155

Patent No. 6117654

GENERAL INFORMATION:

APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 11.4%; Score 57.2; DB 3; Length 152331;

Best Local Similarity 46.6%; Pred. No. 0.0016;

Matches 197; Conservative 0; Mismatches 224; Indels 2; Gaps 1;

QY 50 GCGCGCGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 109
DB 21831 GGTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 21890
QY 110 GCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 169
DB 21891 CCG 21950
QY 170 CCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 229
DB 21951 CCG 22010
QY 230 GCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 289
DB 22011 GCG 22070
QY 290 CAGCG 349
DB 22071 ACAGCGACCG 22128
QY 350 GCGCGCATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 409
DB 22129 CCG 22188
QY 410 CCG 469
DB 22189 CCG 22248
QY 470 ACC 472
DB 22249 CCG 22251

RESULT 12
US-09-165-264-8/c
Sequence 8, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinyagamoorthy, Thuralayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44747
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14


```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

```

Query Match	10.4%;	Score 52.2;	DB 4;	Length 319;
Best Local Similarity	48.2%;	Pred. No. 0.011;		
Matches 147; Conservative	0;	Mismatches 158;	Indels 0;	Gaps 0

[illegible]

Qy	457	CCGCC	461
Db	12	CCCCC	8

```

RESULT 13
US-09-165-264-13/c
Sequence 13, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinayagamoorthy, Thuralayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44747
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

```

Query Match	10.0%;	Score 50;	DB 4;	Length 320;
Best Local Similarity	47.58;	Pred. No. 0.029;		
Matches 149;	Conservative	0;	Mismatches 165;	Indels 0;
			Gaps	0;

QY 148 GCGGGGCCCCGGGGTCTCGGAGTCCGGCCGGCTGTGTCGGCGAGGAGACCCACCCGGCT 207
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Ddb 319 GGCATGACCTCTGCTCCACGTCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 260
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 QY 208 GCGCGCCGATGCTTGCAGTGGGGCCCGCCATGAGACGAGATTACCCGACATGAACCCC 267
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Ddb 259 CCGCCC 200
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 QY 268 GCGCGGGCGGAGACCTCTGTACAGCCCGCGCCCTGTAGAGACGCGCATGCTGACACGCC 327
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Ddb 199 CC 140
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY	328	CCTACTGGAACACTTTCGCGTGCCGCATTTACCCTGTGCTTCTCCAGGAGAACGCCGCT	387
Dd	139	CCC	80
QY	388	TCAATGAGTCCGCTCCTCTCTGGGACACAGCCCTGTGCCAAGCACGATATGCCCG	447
Dd	79	CCC	20
QY	448	TGGCCACCGCCTTC	461
Dd	19	CCCCCCCCCCCCCCC	6

QY	448	TGGCACC	GCTCC	461
Db	19	CCCCCCCC	CCCC	6

RESULT 14
US-08-718-388-5/c
; Sequence 5, Application US/08718388
; Patent No. 6271362

APPLICANT: MORIKAWA, MORISU
 APPLICANT: HARADA, NAOKI
 TITLE OF INVENTION: GENE ENCODING IGG Fc REGION-BINDING
 TITLE OF INVENTION: PROTEIN
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
 STREET: PO BOX 747
 CITY: FALLS CHURCH
 STATE: VA
 COUNTRY: USA
 ZIP: 22040-0747
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/718,388

```

? CLASSIFICATION: 536
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME: MURPHY JR, GERALD M
? REGISTRATION NUMBER: 28,977
? REFERENCE/DOCKET NUMBER: 0230-1111
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: (703) 205-8000
? TELEFAX: (703) 205-8050
? INFORMATION FOR SEQ ID NO: 5:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 3661 base pairs
?
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? MOLECULE TYPE: CDNA
?
? OS-08-718-388-5

```

Query Match	Score	DB	Length
9.8%	49.2	4	3661

	Matches	168;	Conservative	0;	Mismatches	198;	Indels	0;	Gaps	0;
OY	54	CCCCGGGCTCCGCCCGCCGGACCCGGGCCCGAGATCATGATCTCTCCGCACCGCCG	113							
Db	2805	CAGAAGCCTTCAGAGATGATGCTTGAACCTTGCGAGCGTCGTACACAAGCACCTCTGAAATAC	2746							
OY	114	CCACCAACGAGACGAGAACCCAGATAGACGGCCCGCGCGCCCGGCTGTGGAGTCCGGC	173							
Db	2745	TGGCGGGGGCGGCACAAAGGCCGTGGCAAAGGCCGCAAGCGGCCGTTGGTGGCGAATTACG	2686							
OY	174	CAGCTGCTGCCGCGCGAGAACCCACCCCGCTGCGCCCGCATGTCTTGACATGGGGGCC	233							
Db	2685	CCGACAGGCGTCCGGGCGCCGGAAGAGACTCTTTGCTGTGGGGGTGCACAGGGACGGGCAT	2626							
OY	234	GCCATGACACGGATTACCCGACAGATGAACCCC CGCGCGGGCAGCTCTGTACGC	293							

DB 2625 GGGTTGGACACATTTCCCGCAGCCCTGGGGCGCCGCCACCTGCCATCCGGGGGCTTC 2566
OY 294 CCGCCGGCCCTGACAGAGGCCATGCTGCATGCCCCCTACTGGAACCTTCTGCTGCCG 353
DB 2565 CCGCCACCGCGCTTCAGTGTCTCTGCGGGGCTCTGTTGTAGTTCGCGATAAGCCACAG 2506
OY 354 CCATACCCCTGCTCTCTGACAGGACAGCCCGCTTCATGAGTCCGCTCTCTCGGC 413
DB 2505 AGAGAGCCCGGTACGCGCGCGACAGCGCAGCGACGAGCTGTCCCGTGAAGCC 2446
OY 414 AGCCAG 419
DB 2445 AGCGAG 2440

RESULT 15

US-08-718-388-6/c
Sequence 6, Application US/08718388

Patent No. 6271362

GENERAL INFORMATION:

APPLICANT: MORIKAWA, MINORU

APPLICANT: HARADA, NAOKI

TITLE OF INVENTION: GENE ENCODING Igg Fc REGION-BINDING

TITLE OF INVENTION: PROTEIN

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

STREET: PO BOX 747

CITY: FALLS CHURCH

STATE: VA

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/718,388

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR, GERALD M

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 0230-111

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 7824 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 21..7802

US-08-718-388-6

Query Match

Best Local Similarity 9.8%; Score 49.2; DB 4; Length 7824;

Matches 168; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

OY 54 CCGCGGCTCCCGCGCCCGCGGACCCCGGCGCCGAGATCATGATGCTGCAGCCACCGCGC 113
DB 6940 CAGAGGCTCTCCAGATGGCCCTGAACCTTGAGGCGCTCAGCAAGCCCTGGAAGTAC 6881
OY 114 CCACACGAGGAGAGAGCCAGATAGAGCCCGCGGCGCCCGGCTCTGAGTCCCGC 173
DB 6880 TCGCGGCGGCGGACAGAGCCGTGGCAGGGCGCCAGCGGCGCTGGTGGGAGATCAG 6821

OY 174 CGCTGTGCCCCGCGAGAGACCCCGCCCTGCGCCCGGATGCTTGAGTGGGGCC 233
DB 6820 CCGAGGCGTCCGGGCGCGGAGACCTCTGCTGTGGGTTGCACGGCGAGCGGCGAT 6761
OY 234 GCCATGACAGGAGATTACCGCAGCATGAACCCCGCGGCGGCGACCTCTCTACAGC 293
DB 6760 GGTGTGACACACATTCCTCCGCGAGCCCTGGGCGCGCCGACCTGCGCATTCGGGGCTTC 6701
OY 294 CCGCGGCGCTGACAGAGCGCATGCTGCATGCCCCCTACTGGAACACTTCTGCTGCGC 353
DB 6700 CCGCCACCTGCTTCAGTGTCTGTGGGGGCTCCTGTTGTAGTTCGCGCATTAAGCCACAG 6641
OY 354 CCATACCCCTGCTCTCTGACAGGACAGCCCGCTTCATGAGCTCCGCTCTCTGCGC 413
DB 6640 AGAGAGCCCGGTACGCGCGCGCGCAGCGCAGGCGCAGAGATGTCTCCCTGGAAGCC 6581
OY 414 AGCCAG 419
DB 6580 AGCGAG 6575

Search completed: October 10, 2002, 19:27:34
Job time : 104.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds

(without alignments)
9699,805 Million cell updates/sec

Title: US-09-489-101a-13_COPY_1_500

Sequence: 1 gccgcaaacccggaagtgag.....gcgacttgcacgactcc 500

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estha:*
2: em_esthum:*
3: em_estcin:*
4: em_estmov:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_ges:*
13: em_ges_hum:*
14: em_ges_inv:*
15: em_ges_pln:*
16: em_ges_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	460	92.0	483	10	BML48224 TCAP1098
2	396.2	79.2	402	10	BE241632 TCAP106
3	178	35.6	712	10	BE409724 601301584
4	121	24.2	635	9	BB664186
5	116.2	23.2	695	10	BE952238
6	112.8	22.6	131	10	BE242275 TCAP1E15
7	112.8	21.3	2658	11	BC019206 Mus muscu
8	106.6	21.3	635	12	AG084012 Pan trogl
9	79.2	15.8	776	12	CNS010RY
10	78.4	15.7	951	9	AL543262
11	77.4	15.5	612	10	AL543262
12	75	15.0	274	9	AA948370
13	74.8	15.0	834	10	AA948370
14	73.2	14.6	855	12	AA948370
15	72.8	14.6	798	10	AA948370
16	72.8	14.6	848	12	AA948370
17	72.4	14.5	559	9	AA948370

C 18	72.4	14.5	932	12	CNS00720	AL066742 Drosophila
C 19	71.6	14.3	810	12	AG060267	AG060267 Pan trogl
C 20	71	14.2	741	12	A2194927	A2194927 SP_1029_A
C 21	70.6	14.1	576	10	BE786264	BE786264 SEAMC006
C 22	70.4	14.1	927	10	BE786316	BE786316 SEAMC006
C 23	70.2	14.0	1339	12	AG137991	AG137991 Pan trogl
C 24	70	14.0	925	12	CNS0091P	AL053013 Drosophila
C 25	70	14.0	1023	12	AG128304	AG128304 Pan trogl
C 26	70	14.0	1101	12	CNS0153F	AL104949 Drosophila
C 27	69.8	14.0	918	10	BE809598	BE809598 mgct001xb
C 28	69.6	13.9	821	10	BE809582	BE809582 mgct001xb
C 29	69.6	13.9	831	10	BE952198	BE952198 HYSMEM000
C 30	69.6	13.9	1080	12	AG093056	AG093056 HS_4832_A
C 31	69.4	13.9	821	12	AG093474	AG093474 HS_4832_A
C 32	69.4	13.9	1128	10	BM477735	BM477735 ACENICOURT
C 33	69.2	13.8	641	9	AL137868	AL137868 qv13b02.x
C 34	69.2	13.8	925	10	BE441241	BE441241 GA_Ea001
C 35	69	13.8	908	12	CNS006B4	AL064031 Drosophila
C 36	68.6	13.7	643	12	AG123077	AG123077 Pan trogl
C 37	68.4	13.7	827	12	A2183795	A2183795 SP_1002_A
C 38	68.2	13.6	613	12	AG043036	AG043036 Pan trogl
C 39	68.2	13.6	670	12	AG126228	AG126228 Pan trogl
C 40	68.2	13.6	836	12	A2185303	A2185303 SP_1005_A
C 41	68.2	13.6	863	12	AG1332157	AG1332157 Pan trogl
C 42	68.2	13.6	864	12	AG060559	AG060559 Pan trogl
C 43	68.2	13.6	880	12	AG042453	AG042453 HS_5383_B
C 44	68.2	13.6	925	9	AL581447	AL581447
C 45	68.2	13.6	925	12	CNS0091P	AL053013 Drosophila

ALIGNMENTS

RESULT 1
BML48224
LOCUS
DEFINITION
TCAP109825 Pediatric acute myelogenous leukemia cell (FAB M1)
Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAA9825, mRNA
sequence.
ACCESSION
BML48224
VERSION
BML48224.1
KEYWORDS
GT:17168744
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (Bases 1 to 483)
Wang, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman, F.R., Jr.,
Gunaratne, P.H., Muzny, D., Bouck, J., Gibbs, R.A., and Margolin, J.F.
Pediatric leukemia cDNA Sequencing Project (2001)
unpublished (2001)
CONTACT: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@xccc.org
Seq primer: M13 primer:
Location/Qualifiers
1. 483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAA9825"
/clone_lib="Pediatric acute myelogenous leukemia cell (FAB M1)
Baylor-HGSC project-TCAA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"
/note="Vector: lambda PSB; site_1: BamHI; site_2: EcoRI;
First strand cDNA was primed with an anchored

LOCUS BE409724 712 bp mRNA linear EST 21-JUL-2000
 DEFINITION 601301584F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3636141 5',
 ACCSSION BE409724 mRNA sequence.
 VERSION BE409724.1 GI:9346174
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 712)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ULNL at: image.jnl.gov
 Plate: LICM33 row: 9 column: 22
 High quality sequence start: 6
 High quality sequence stop: 686.
 Location/Qualifiers
 1..712
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3636141"
 /clone_lib="NIH MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Placenta; Vector: pOT7; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(6). Size selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 110 a 235 c 240 g 127 t
 ORIGIN
 Query Match 35.6%; Score 178; DB 10; Length 712;
 Best local similarity 90.2%; Pred. No. 1.6e-20;
 Matches 202; Conservative 0; Mismatches 20; Indels 2; Gaps 1;
 QY 88 AGATCATGATGCTGGCGGCGGACGACGAGGAGGAGCCGATGAGCGCCC 147
 DB 453 AGATCATGATGCTGGCGGCGGACGACGAGGAGGAGCCGATGAGCGCCC 512
 QY 148 GGGCGGCGGCGGCTCTGGAGTCCGCGGCTGTCGCCGCGGAGACCCCGGCT 207
 DB 513 GGGCGGCGGCGGCTCTGGAGTCCGCGGCTGTCGCCGCGGAGACCCCGGCT 572
 QY 208 GCGCGCGGAGTCTGAGTGGGCGGCGGATGAGGAGGAGTACCGGAGGAGCCG 267
 DB 573 GCGCGCGGAGTCTGAGTGGGCGGCGGATGAGGAGGAGTACCGGAGGAGCCG 632
 QY 268 GCGCGGCGGAGGCTCTGAGTGGGCGGCGGCGGCGGAGGAGGAGGAGGAGG 311
 DB 633 GCG--GCGGCGGAGGCTCTGAGTGGGCGGCGGCGGAGGAGGAGGAGGAGG 674

RESULT 4
 LOCUS BB664186 635 bp mRNA linear EST 26-OCT-2001
 DEFINITION BB664186 RIKEN full-length enriched, 0 day neonate lung Mus
 ACCESSION BB664186
 VERSION BB664186.1 GI:16497940
 KEYWORDS EST.

SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 635)
 AUTHORS Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, U., Kono, H., Kouda
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
 D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T.,
 Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 TITLE Unpublished (2001)
 JOURNAL Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E.,
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1737-1771 (2000)
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1..635
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="E030029A20"
 /clone_lib="RIKEN full-length enriched, 0 day neonate
 lung"
 /tissue_type="lung"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site_1: Sali; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGGAGAGAGCGCGCGGACGAGCTTTTCTTTTCTTTTCTTTT 3']. cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGGAGAGAGTCTCGAGTAAATTAATTCATCCCGCCCCCGCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 pluescript KS(+) after bulk excision from lambda f1c1."

BASE COUNT 121 a 204 c 182 g 126 t 2 others
ORIGIN

Query Match 24.2% Score 121; DB 9; Length 635;
Best Local Similarity 60.48; Pred No. 4.4e-11;

Matches 299; Conservative 0; Mismatches 145; Indels 51; Gaps 4;

OY 1 GCCCGAAACCCGAGTGAAGCGGCGGCGGCTCGAGGCTCGAGAAACGCGCGG 60
DB 6 GCCCGAAACCCGAGTGAAGCGGCGGCGGCTCGAGGCTCGAGAAACGCGCGG 63
OY 61 CTCGCGCGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 120
DB 64 --GGGCGTGAAGGATGAGTGTGAGATATGACATTAACACTGCGCGTGCACAC 120
OY 121 GGAGCGAAAGCCAGTGAAGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAG 180
DB 121 AGAGTGAAGGATGAGTGTGAGATATGACATTAACACTGCGCGTGCACAC 175
OY 181 TGCCCGCGGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 240
DB 176 TGCGCT-----GATGCTGCGAGTGAAGCGGCGGCTCGAGGCTCGAGGCT 206
OY 241 ACAGGATTAACCGGAGTGAAGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCT 300
DB 207 ATGGGAGTTCCTCGGAGTGAAGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAG 256
OY 301 CCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 360
DB 257 --CTACAGAAACCGGAGTGAAGCGGCGGCGGCGGCTCGAGGCTCGAGGCT 314
OY 361 CTCGCTTCGAGGAGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 420
DB 315 CCACCTTCGAGGAGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 374
OY 421 CCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 480
DB 375 CTGCGCTGAGGAGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 434
OY 481 GCGACTTTCGAGGAGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 495
DB 435 GCGACTTTCGAGGAGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 449

RESULT 5 BE952238 695 bp mRNA linear EST 04-OCT-2000
LOCUS UI-M-CD0-ayk-d-03-0-UI-s1 NIH BMAP Ret2 Mus musculus cDNA clone
DEFINITION UI-M-CD0-ayk-d-03-0-UI-s1, mRNA sequence.
ACCESSION BE952238
VERSION BE952238.1 GI:10592495
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 695)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@nhi.nih.gov
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq. Primer: M13 Forward
POLY-A-MO. Location/Qualifiers

FEATURES
Source

1.695
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="UI-M-CD0-ayk-d-03-0-UI"
/clone_id="NIH_BMAP_Ret2"
/dev_stage="1 day"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_Ret2 library is derived from mouse neonatal
retina tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at brainest.eng.utoronto.edu.
TAG_SEQ=None found"

BASE COUNT 145 a 234 c 173 g 143 t
ORIGIN

Query Match 23.2% Score 116.2; DB 10; Length 695;
Best Local Similarity 64.98; Pred. No. 2.8e-10;
Matches 194; Conservative 0; Mismatches 93; Indels 12; Gaps 1;

OY 202 CCGCGCGCGCGCGGAGTGAAGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCT 261
DB 140 CTCGCGAGGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 199
OY 262 AACCGCGCGGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 321
DB 200 AGCTCCAGCGGCGGCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 247
OY 322 ACTGCCCTTCTGAGAACCTTCTGCGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 381
DB 248 ACTGCCCTTCTGAGAACCTTCTGCGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 307
OY 382 GCGCGTTCATGAGCTCGGCTCGGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 441
DB 308 AACGATTCGTTGGTTCACACCCATTCCTTGTGGCGGAGTTCCTCGTGAACACCTACCTAC 367
OY 442 CCGCGCGCGCGCGGAGTGAAGCGGCGGCGGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 500
DB 368 CCACGACGACGACGCTCCCGGCTTCCTTCGCAAGAGGAGGAGCTTCCTCGAGGAGCC 426

RESULT 6 BE242275 131 bp mRNA linear EST 03-OCT-2001
LOCUS TCAAP1545 Pediatric acute myelogenous leukemia cell (FAB M1)
DEFINITION Baylor-HGSC Project-TCAA Homo sapiens cDNA clone TCAAP1545, mRNA
sequence.
ACCESSION BE242275
VERSION BE242275.1 GI:9094002
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 131)
AUTHORS Wei, Y., Tsang, Y.T.M., Mel, G., Ku, J.M., Ali-Osman, J.R., Muzny, D.
Bouck, J., Gibbs, R.A. and Margolin, J.F.
TITLE Pediatric Leukemia cDNA Sequencing Project
JOURNAL Unpublished (2000)
COMMENT Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038

Email: clones@cycc.org
 Curation: Carninci, P. and Hayashizaki, Y. High efficiency
 full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Seq primer: M13 primer.

FEATURES

Location/Qualifiers
 1..131
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="TCAP1545"
 /clone_1lb="pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HESC project-TCAP"
 /sex="male"
 /tissue_type="leukopheresis"
 /cell_type="myeloid cell"
 /dev_stage="pediatric 6 years"
 /lab_host="DH10B"
 /note="Vector: lambda psb. Site_1: BamHI; Site_2: EcoRI; First strand cDNA was primed with an anchored XhoI-oligo(dT) primer [5'GAGAGCTGAGCGCGCGAGAGAG(T)VN 3'; V-A,C,G; N-A,C,G,T] and then dg tailed. Second strand was primed with a BamHI-dC primer [5'AGAGAGCTGAGCGCGCGCGCGCAATTAATTAAT(C) 3']. Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda psb vector. Library was constructed by Wei Yu at RIKEN of Japan (Carninci P., Westover A., Nishiyama Y., Ohsumi T., Itoh M., Nagaoka S., Sasaki, Okazaki Y., Muramatsu M., Schneider C., Hayashizaki Y., High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb 28, 1997)".

BASE COUNT 19 a 47 c 49 g 15 t 1 others
 ORIGIN

Query Match 22.6% Score 112.8; DB 10; Length 131;
 Best Local Similarity 93.6% Pred. No. 9.3e-10;
 Matches 117; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 22 GGGCGAGCTGCTGCGAGGAGAAAGCGCGCGCTCCGCGCGCGAGCCG 81
 |||||
 Db 2 GGGCGAGCTGCTGCGAGGAGAAAGCGCGCGCTCCGCGCGCGAGCCG 61
 |||||
 QY 82 GGGCGAGATCATGATGCTGCGCGCGCGCGCGAGCGAGCGAGCCAGATGA 141
 |||||
 Db 62 GGGCGAGATCATGATGCTGCGCGCGCGCGCGAGCGAGCGAGCCAGATGA 121
 |||||
 QY 142 CGGCC 146
 |||||
 Db 122 CGGCC 126

RESULT 7
 LOCUS BC019206 2658 bp mRNA linear HTC 11-DEC-2001
 DEFINITION Mus musculus, similar to KIAA0963 protein, clone IMAGE:5052084,
 mRNA.

ACCESSION BC019206
 VERSION BC019206.1 GI:17512508
 KEYWORDS HTC.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
 1 (bases 1 to 2658)
 AUTHORS Strausberg R.
 JOURNAL Direct Submissi
 Submitted (07-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,
 USA

REMARK NIH-MGC project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nih.gov
 Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, O.L., Mastaglio, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stantrid, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H., and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: FRAX Plate, 39 Row; m Column; 8
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein
 This clone has the following problem: incomplete processing.

FEATURES

Location/Qualifiers
 1..2658
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5052084"
 /tissue_type="liver, normal, 5 month old male mouse."
 /clone_1lb="M1 CGAP_119"
 /lab_host="DH10B"
 /note="Vector: pCMV-Sport6"

BASE COUNT 600 a 761 c 696 g 601 t
 ORIGIN

Query Match 22.6% Score 112.8; DB 11; Length 2658;
 Best Local Similarity 64.6% Pred. No. 1.1e-09;
 Matches 190; Conservative 0; Mismatches 92; Indels 12; Gaps 1;

QY 202 CGCGCTGCGCGAGATGCTTGCATGCGGCGCGCGATGAGACAGATTAACCGACATG 261
 |||||
 Db 139 CTTCCATGCTGCGAGATGCTGAGTGAAGACCCAGATGAGTGGAGCTTCCATGATG 198
 |||||
 QY 262 AACCCCGCGCGGCGGAGCAGCTCTGTACAGCCCGCGCGCGCGCGCGATGCTGC 321
 |||||
 Db 199 AGCTCCACCCCTCTGAGGTGATCCAG-----CTACGAACCGACATCTGC 246
 |||||
 QY 322 ACTGCCCTTACTGGAACACTTCTGCTGCGCGCGCGATGCTTCTTCCAGACGACCC 381
 |||||
 Db 247 ACTGCTTGTGTGGGAGATTTCTCACCCTTGTACCCAGACCTTCTCAGTGAAGAAC 306
 |||||
 QY 382 GCGGCTCATGAGCTTCCGCTTCTGCGCGAGCAGCGCGCGCGCGAGCAGCAGCTATG 441
 |||||
 Db 307 AACAGTTCGTGGTTCACACCATCTTGTGTGCGAGTGTGCGTGTGCGTGAAGACAGTAC 366
 |||||
 QY 442 CCCCCGCGCGAGCAGCTTCCAGCTTGCACCAAGACCTGCGACTTGTCTCAGG 495
 |||||
 Db 367 CCACCAAGCCAGCTTCCAGCTTCTTCTCAGAGCAGCAGCTTCTCAGG 420
 |||||

RESULT 8
 LOCUS AG084012 635 bp DNA linear GSS 03-NOV-2001
 DEFINITION Pan troglodytes DNA, clone: PTB-081K02.R, genomic survey sequence.
 AG084012

ACCESSION AG084012
 VERSION AG084012.1 GI:16635814
 KEYWORDS GSS: GSS (genome survey sequence).
 SOURCE pan troglodytes male lymphoblast DNA, clone: PTB-081K02.R.
 ORGANISM Pan troglodytes

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 1 (sites)

Fri Oct 11 09:30:42 2002

us-09-489-101a-13_copy_1_500.rst

Page 6

AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	BAC end sequences of library PTB
JOURNAL	Unpublished
REFERENCE	2 (Phases 1 to 635)
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Suenho-chou, Tsukuba-shi, Ibaraki, 305-8565, Japan tel: 81-45-503-9111, fax: 81-45-503-9170 e-mail: schimpess@gsc.riken.go.jp, URL: http://hgp.riken.go.jp/ Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
COMMENT	PRIMERS Sequencing: M13Rev
FEATURES	LIBRARY Vector : pKS145 R.Site 1 : SacI R.Site 2 : SacI. Location/Qualifiers 1. 635 /organism="Pan troglodytes" /db_xref="taxon:9598" /clone="PTB-081K02.R" /sex="male" /cell_type="lymphoblast" /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT	104 a 236 c 189 g 105 t 1 others
ORIGIN	
Query Match	21.3% Score 106.6; DB 12; Length 635;
Best Local Similarity	95.6%; Pred. No. 1,1e-08;
Matches 109; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
OY	382 GCCCGTCATGACCTCCGCTCTCTTCCTTCGCGACGACCCCTGCGCCAGACACGCTTANG 441
DB	23 GCGTCTCGAGACCTCGGCTCTCTTCCTTCGCGAGCGACCCCTGCGCCAGACACGCTTANG 82
OY	442 CCCCCGTGGCCACCGCCTCCACGCTTGCACCAAGACCTGCGACTTGGCTGACG 495
DB	83 CCCCCGTGGCCACCGCCTCCACGCTTGCACCAAGACCTGCGACTTGGCTGACG 136
RESULT 9	
CNS010RY	
LOCUS	CNS010RY 776 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC
LOCATION	BACN04P24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL0993352
VERSION	AL0993352.1 GI:5610963
SOURCE	GSS.
ORGANISM	fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (Bases 1 to 776)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC end sequence was carried out as part of a collaboration with the European Drosophila genome project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC Project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector
COMMENT	

[illegible]

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@life.com URL: http://fulllength.invitrogen.com"

BASE COUNT 122 a 492 c 154 g 118 t 65 others

ORIGIN

Query Match 15.7%; Score 78.4; DB 9; Length 951;

Best Local Similarity 44.2%; Pred. No. 0.00052;

Matches 184; Conservative 28; Mismatches 204; Indels 0; Gaps 0;

57 CGGGCTCCGCGCCGCGGAGCCGAGATCATGATGTCGCGCCAGCCGCA 116
 6 CGCGTGAAGCTTCGACCTACCTCCGCGCCGCGCTCCCTCCCTCCCTCC 65
 117 CCAGGAGCGAAGAGCAGATAGACGCGCGCGCGCGCGCGCGCGCGCG 176
 66 CCGCTCCG 125
 177 CTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 236
 126 CCG 185
 237 ATGAGCAGAGGATATACCGAGCATGACCGCGCGCGCGCGCGCGCG 296
 186 SCGGCG 245
 297 CG 356
 246 CCG 305
 357 TACCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
 306 MCACCG 365
 417 CAGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 472
 366 CCG 421

RESULT 11
 B1149078 612 bp mRNA linear EST 05-JUL-2001
 LOCUS B1149078
 DEFINITION 60291075F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5052084 5',
 mRNA sequence.
 ACCESSION B1149078
 VERSION B1149078.1 GI:14609079
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 612)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: L14M1141 row: m column: 13
 High quality sequence start: 14
 High quality sequence stop: 581.

FEATURES

source

Location/Qualifiers
 1. 612
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5052084"
 /clone="1D" NCI_CGAP_L19
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 122 a 199 c 165 g 126 t

ORIGIN

Query Match

Best Local Similarity 61.8%; Score 77.4; DB 10; Length 612;

Matches 181; Conservative 0; Mismatches 96; Indels 16; Gaps 3;

202 CCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
 150 CCTCAGCTTCCCTGATGCTGAGTGAACCGACACGATGAGTGGACTTCC 209
 262 AACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 321
 210 ACCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 257
 322 ACTGCCCTACTGAAACCTTCGCGCGCGCGCGCGCGCGCGCGCGCG 381
 258 ACTGTCCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 317
 382 GCGCGTTC-ATGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
 318 AACAGTCTGTTGGTTCACCCATCTCTTGCGCGCGCGCGCGCGCGCG 377
 441 GCCCG 490
 378 CCATCAGACAGACAGTACACAGCTTCTTCACAAAGCAGCAGATTC 430

RESULT 12
 AA948370 274 bp mRNA linear EST 04-MAY-1998
 LOCUS AA948370
 DEFINITION og46f04.s1 NCI_CGAP_K1d5 Homo sapiens cDNA clone IMAGE:1589407 3',
 mRNA sequence.
 ACCESSION AA948370
 VERSION AA948370.1 GI:3109623
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 274)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bdb.lnl.gov/bdrp/image/image.html
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 228.

FEATURES

source

Location/Qualifiers
 1. 274
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES	Location/Qualifiers
source	1. .855

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 84.9333 Seconds

(WARNING: alignments 10107.424 Million cell updates/sec)

Title: US-09-489-101a-13_COPY_1_500

Sequence: 1 gcccggaaccgcgaagtgg.....gcgacttgcctcgaactcc 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: N.Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	500	100.0	4877	22	AA011120 Human small cell 1
2	500	100.0	4924	21	AA075998 Human OREX1553
3	173	34.6	343	20	AA07103 EST clone BH1. H
4	110.8	22.2	114	21	AA015646 Human secreted pro
5	104.6	20.9	1553	22	AA037392 Human musculoskele
6	65	13.0	1000	21	AA02484 Human colon cancer
7	60	12.0	799	19	AA05831 Nucleotide sequenc
8	60	12.0	1218	21	AA02488 Human colon cancer
9	60	12.0	1925	20	AA090924 Epstein Barr Virus

10	60	12.0	1926	21	AA050254 Epstein Barr Virus
11	60	12.0	1926	22	AA082902 EBV tethering prot
12	60	12.0	2580	21	AA075454 Nucleotide sequenc
13	60	12.0	5452	20	AA090923 Anti-sense strand
14	60	12.0	8705	20	AA023778 Vector pShuttle DN
15	60	12.0	9600	19	AA021683 Nucleotide sequenc
16	60	12.0	10280	20	AA022248 Nucleotide sequenc
17	60	12.0	10286	14	AA031348 Plasmid pCISBON f
18	60	12.0	10536	20	AA015650 Nucleotide sequenc
19	60	12.0	10596	20	AA015650 DNA clone pCK C1
20	60	12.0	16080	21	AA059553 Human gene express
21	59.4	11.9	1337	20	AA017263 Human IL-1ra BAC c
22	58.4	11.7	2561	22	AA028500 Human adenosine A1
23	57.8	11.6	114955	20	AA053491 Human IL-1ra BAC c
24	57.2	11.4	3198	20	AA02974 Genomic DNA encodi
25	55.6	11.1	795	19	AA055830 Human genome from
26	53.8	10.8	7720	21	AA053800 Human adenosine A1
27	53	10.6	114793	22	AA080215 Human low density
28	52.2	10.4	114955	20	AA053491 Human IL-1ra BAC c
29	51.2	10.2	1614	22	AA028459 Human low density
30	51.2	10.2	12425	22	AA028459 Human low density
31	50.8	10.2	12427	21	AA028459 Human low density
32	50.8	10.2	12427	21	AA028459 Human low density
33	50.2	10.0	2198	24	AB099430 Mouse ischaemic co
34	50.2	10.0	320	21	AA038185 DNA encoding novel
35	49.4	9.9	3492	23	AA083651 Active clone V11 o
36	49.2	9.8	3661	16	AA063081 DNA encoding novel
37	49.2	9.8	7824	16	AA063073 DNA encoding novel
38	49.2	9.8	8875	23	AA08401 IgG-Fc binding pro
39	49.2	9.8	16382	16	AA063074 DNA encoding novel
40	49.2	9.8	16421	21	AA086623 Human colon cancer
41	48.8	9.8	1459	21	AA00528 Human encoding novel
42	48.8	9.8	1459	23	AA08406 Human encoding novel
43	48.8	9.8	1459	23	AA08406 Human encoding novel
44	48.8	9.8	1459	23	AA08406 Human encoding novel
45	48.8	9.8	1459	23	AA08406 Human encoding novel

ALIGNMENTS

RESULT 1	
AD011120	standard: DNA: 4877 BP.
ID	AD011120
AC	AD011120
XX	24-SEP-2001 (first entry)
XX	Human small cell lung cancer associated gene, K1A0963.
XX	Human: small cell lung cancer; therapy: hCAP; nucleic acid: NA;
KW	melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW	leiomyosarcoma; synovial sarcoma; cytostatic; ds.
XX	Homo sapiens.
OS	Homo sapiens.
XX	WO200153349-A2.
XX	26-JUL-2001.
PD	19-JAN-2001; 2001MO-US02015.
XX	21-JAN-2001; 2000US-0489101.
XX	(LUDWIG) LUDWIG INST CANCER RES.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.
PA	(CORR) CORNELL RES FOUND INC.
PI	Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX	WPI: 2001-457597/49.

08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-451937/48.
 Isolated polypeptide for treating, preventing and/or prognosing
 disorders related to the musculoskeletal system including
 musculoskeletal cancers and also for testing and detection e.g.
 diagnosis -
 Example 2: SEQ ID NO 3757; 781bp + Sequence Listing: English.
 The invention relates to novel genes (AAL34669-AAL37666) and proteins
 (AAB03087-AAB04109) associated with the musculoskeletal system useful
 for preventing, treating or ameliorating medical conditions e.g. by
 protein or gene therapy. The genes are isolated from a range of human
 tissues disclosed in the specification. The nucleic acids, proteins,
 antibodies and (ant)agonists are useful in the diagnosis, treatment
 and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 other cancers of the adrenal gland, bone, bone marrow, breast,
 gastrointestinal tract, liver, lung, or urogenital; (b) immune
 disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 and (f) infectious diseases such as viral, bacterial, fungal and
 parasitic infections.
 Note: The sequence data for this patent did not form part of the
 printed specification, but was obtained in electronic format directly
 from WPI at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1553 BP; 294 A; 461 C; 563 G; 235 T; 0 other;
 Query Match 20.9%; Score 104.6; DB 22; Length 1553;
 Best Local Similarity 74.9%; Pred. No. 5.8e-12;
 Matches 131; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GCCCGAAACCCGGAAGTGAAGCGCGGACGCTGCGAGCTCGAGAAACAGCGCGCGG 60
 DB 921 GCCCGAAACCCGGAAGTGAAGCGCGGACGCTGCGAGCTCGAGAAACAGCGCGCGG 862
 QY 61 CTCG 120
 DB 861 CTCG 802
 QY 121 GAGCGAGAGAGCCGATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 175
 DB 801 TCACCG 747
 RESULT 6
 ID AAA02484 standard; cDNA; 1000 BP.
 AC AAA02484;
 DT 19-MAY-2000 (first entry)
 XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2475.
 DE Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX Homo sapiens.
 OS WO9558675-A2.
 PN 18-NOV-1999.
 PD 13-MAY-1999; 99MO-US10602.
 PF 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085337.
 PR 15-MAY-1998; 98US-0085696.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Leshkowitz D, Kila D, Garcia V, Jones LM, Staehle-Crain B;
 DR WPI: 2000-126369/11.
 XX Polynucleotide library used to determine cancerous states of mammalian
 cells -
 Claim 1; Page 994; 1097pp: English.
 AA000010 to AA02716 represent polynucleotides isolated from cDNA
 libraries constructed from human colon cancer cell lines. The present
 invention also describes a method of detecting differentially expressed
 genes correlated with a cancerous state of a mammalian cell, comprising
 detecting at least one differentially expressed gene product in a test
 sample derived from a cell suspected of being cancerous, where detection
 of the differentially expressed gene product is correlated with a
 cancerous state of the cell from which the test sample was derived.
 The polynucleotide sequences can be used in a method for detecting
 differentially expressed genes correlated with a cancerous state of a


```

DT 19-MAY-2000 (first entry)
XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2479.
DE
XX Human: colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
XX oestrogen receptor-negative breast cancer; lung cancer; ss.
OS Homo sapiens.
XX
XX WO958675-A2.
XX
XX 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US10602.
XX
XX 14-MAY-1998; 98US-0085426.
XX 15-MAY-1998; 98US-0085337.
XX 15-MAY-1998; 98US-0085696.
XX 21-OCT-1998; 98US-0105234.
XX 27-OCT-1998; 98US-0105877.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX Lamson G, Dzman R, Cirvenjakov R, Dickson M, Dzmanac S, Labat I;
XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI: 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
XX cells -
XX
XX Claim 1; Page 995-996; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
XX libraries constructed from human colon cancer cell lines. The present
XX invention also describes a method of detecting differentially expressed
XX genes correlated with a cancerous state of a mammalian cell, comprising
XX detecting at least one differentially expressed gene product in a test
XX sample derived from a cell suspected of being cancerous, where detection
XX of the differentially expressed gene product is correlated with a
XX cancerous state of the cell from which the test sample was derived.
XX The polynucleotides sequences can be used in a method for detecting
XX a differentially expressed genes correlated with a cancerous state of a
XX mammalian cell. The polynucleotides can also be used as probes for
XX detecting and mapping related genes. They can be used in diagnosis and
XX prognosis of diseases and disorders (e.g. identification of
XX pre-metastatic or metastatic cancerous states, stages of cancer, or
XX responsiveness of cancer to therapy). This is particularly for breast
XX cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
XX negative breast cancer, lung cancer, and colon cancer.
XX
XX Sequence 1218 BP; 9 A; 31 C; 494 G; 37 T; 647 other;
XX
XX Query Match 12.0%; Score 60; DB 21; Length 1218;
XX Best Local Similarity 27.6%; Pred. No. 0.0029;
XX Matches 123; Conservative 0; Mismatches 323; Indels 0; Gaps 0;
XX
XX 55 CGGGGGCTCGCGCGCGGCGGCGGCGGATCATGATGCTGCGCGCACCGCGCG 114
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 972 CACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 115 CACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 174
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 912 CACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 853
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 175 GCCTGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 234
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 852 CCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 793
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 235 CCATGACAGGAGATTAACCCGACGATGAAACCCCGCGGCGGAGGCTCTGTACAGCC 294
DB 792 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 733
QY 295 CGCGCGCGCTGACAGAGGCGCATGCTGCACTGCCCTACTGAGACACTTCTGCTGCCGC 354
DB 732 CCCCCCCCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 673
QY 355 CATACCTGCTTCTCCAGAGGAGCGCGCGCTTCATGAGTCTCCGCTCTCCGCGCA 414
DB 672 CNNCCCGCNCNNCCCGCCGCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 613
QY 415 GCCAGCGCTGCCAGACACAGCATATGCGCGGACCGCTCCAGCTTGCCACCA 474
DB 612 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 553
QY 475 AGACCTGCGACTTGTCTCAGAGACTCC 500
DB 552 NCCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 527

RESULT 9
AAK90924/c
ID AAK90924 standard; DNA; 1925 BP.
XX
XX AAK90924:
XX
XX 17-JAN-2000 (first entry)
XX
XX Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.
XX
XX Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pcMVEBNA; EBNA 1;
XX episome; transfection; origin of replication; EBV oriP; receptor;
XX eucaryotic host cell; recombinant cell line; ion channel; gene therapy;
XX multiple gene expression; transporter protein; transcription factor;
XX adhesion molecule; antisense therapy; gene amplification;
XX cell immortalisation; ds.
XX
XX Epstein-Barr virus.
XX
XX Key Location/Qualifiers
XX CDS 1..1925
XX /tag= a
XX /product= "EBNA 1"
XX /transl_except= (pos:799..800, aa:Gly)
XX /note= "The sequence is described throughout the
XX specification as being 1926 nucleotides long, but a
XX sequence of only 1925 bp has been given in figure 2"
XX
XX WO9947647-A1.
XX
XX 23-SEP-1999.
XX
XX 12-FEB-1999; 99WO-US03307.
XX
XX 18-MAR-1998; 98US-0040961.
XX 06-AUG-1998; 98US-0130114.
XX
XX (PHAR-) PHARMACOPEDIA INC.
XX
XX Dama J BB, Horlick RA, Robbins AK.
XX
XX WPI: 1999-610610/52.
XX P-PSDB; AAY28843.
XX
XX New method for expressing genes from recombinant eukaryotic cells,
XX useful for gene therapy -
XX
XX Claim 24; Fig 2; 86pp; English.
XX
XX The present sequence is a DNA encoding Epstein Barr Virus Nuclear
XX Antigen 1 (EBNA 1), which is obtained from commercially available
XX

```


XX EBV tethering protein EBNA1 encoding DNA.
 XX
 XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
 KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
 KW EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.
 XX
 OS Epstein-Barr virus.
 XX
 FH Location/Qualifiers
 FT 1..1926
 FT CDS /tag=a
 PN WO200125484-A2.
 PD 12-APR-2001.
 PF 29-SEP-2000; 2000WO-US26908.
 XX
 PR 01-OCT-1999; 98US-0410399.
 XX
 XX (UNMI) UNIV MICHIGAN.
 XX
 PA Robertson ES, Colter MA;
 PI WPI: 2001-281736/29.
 DR P-PSDB; AAB63332.
 XX
 PA A composition for use in gene therapy comprises an expression vector
 PI that includes a nucleic acid sequence encoding a nucleic acid binding
 XX protein -
 PS
 PS Disclosure: Fig 9C; 60pp; English.
 XX
 CC The invention provides a composition comprising nucleic acid, histone H1
 CC protein and expression vector operatively encoding a protein suitable
 CC for tethering the nucleic acid to the histone H1 protein, where the
 CC tethering protein is LANA. The composition is useful in aiding the
 CC refection of the viral DNA in the host cell. The viral vector encodes a
 CC protein suitable for tethering DNA to histone H1. Methods for screening of
 CC for compounds which are agonistic or antagonistic for the tethering of
 CC viral proteins to histone H1 and DNA binding sites are useful for
 CC developing the method of viral transfer. The composition has applications
 CC to gene therapy, including the treatment of multiple sclerosis,
 CC Parkinson's disease, Huntington disease and diabetes. The present
 CC sequence represents the nucleotide sequence of the Epstein-Barr virus
 CC (EBV) tethering protein EBNA1.
 CC
 XX
 XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;
 SQ
 Query Match 12.0%; Score 60; DB 22; Length 1926;
 Best Local Similarity 50.6%; Pred. No. 0.003;
 Matches 170; Conservative 0; Mismatches 165; Indels 1; Gaps 1;
 QY 131 GCCAGATAGACGCCCGGGGCGGCGGCTGGAATCCCGCGCGCTGCGCCGCGG 190
 Db 638 GCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 579
 QY 191 AGAGACCCACCG 250
 Db 578 GCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 519
 QY 251 CCGGAGCATGACGCCCGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 310
 Db 518 CCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 459
 QY 311 CGCGATGCTGC-ACGCGCGCTACTGAGAACCTTCTGCGCGCGCGCGCGCGCG 369
 Db 458 CCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 399
 QY 370 CGAGCGACG 429
 Db 398 CCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 339

OY 430 ACACAGCATGATGCCCGCGGAGCCACCGCGCTCCAGCT 465
 Db 338 CCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 303
 RESULT 12
 ID AAA75454/C
 XX AAA75454 standard; DNA; 2580 BP.
 AC
 XX
 AC AAA75454;
 XX
 DE 15-JAN-2001 (first entry)
 XX
 DE Nucleotide sequence of the Epstein Barr nuclear antigen.
 XX
 KW Origin of replication; protein-protein interaction; replication;
 XX two-hybrid system; nuclear antigen; ss.
 XX
 OS Epstein-Barr virus.
 XX
 PN US6114111-A.
 XX
 PD 05-SEP-2000.
 XX
 PF 30-MAR-1998; 98US-0050863.
 XX
 PR 30-MAR-1998; 98US-0050863.
 XX
 PA (RIGE-) RIGEL PHARM INC.
 XX
 PI Luo Y, Payan D, Huang B;
 DR WPI: 2000-593546/56.
 XX
 PT Composition for detecting protein-protein interactions in a mammalian
 PT two hybrid system comprises bait and test vector which consist of
 PT selection gene, vector viral origin of replication and fusion gene -
 XX
 PS Disclosure: Column 17-20; 18pp; English.
 XX
 CC The present sequence represents the Epstein Barr nuclear antigen.
 CC It is used to produce bait vectors of the invention. The specification
 CC describes a compositions and methods for a genetic system of detecting
 CC protein-protein interactions in a mammalian host cell. The system
 CC comprises bait and test, both containing selection genes, and viral
 CC origin of replication. The compositions are useful for detecting an
 CC interaction between a bait protein and a test protein. It is useful in
 CC a mammalian two-hybrid system for detecting protein-protein interactions
 CC in a mammalian host cell.
 CC
 XX
 XX Sequence 2580 BP; 632 A; 512 C; 1054 G; 382 T; 0 other;
 SQ
 Query Match 12.0%; Score 60; DB 21; Length 2580;
 Best Local Similarity 50.6%; Pred. No. 0.003;
 Matches 170; Conservative 0; Mismatches 165; Indels 1; Gaps 1;
 QY 131 GCCAGATAGACGCCCGGGGCGGCGGCTGGAATCCCGCGCGCTGCGCCGCG 190
 Db 1021 GCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 962
 QY 191 AGAGACCCACCG 250
 Db 961 GCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 902
 QY 251 CCGGAGCATGACGCCCGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 310
 Db 901 CCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 842
 QY 311 CGCGATGCTGC-ACGCGCGCTACTGAGAACCTTCTGCGCGCGCGCGCGCGCG 369
 Db 841 CCGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCT 782

Fri Oct 11 09:30:40 2002

us-09-489-101a-13_copy_1_500.png

Job time : 103 secs

Fri Oct 11 09:30:38 2002

us-09-489-101a-12_copy_1_500.rni

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 18.733 Seconds
(Without alignments)
6556.054 Million cell updates/sec

Title: US-09-489-101a-12_COPY_1_500

Sequence: 1 aaagctgtgtcattatattt.....ccctgcgcggagcagaagaag 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Capext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCNUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/Backfiles1.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	31.8	6.4	US-08-287-442-3	Sequence 3, Appl1
2	31.8	6.4	US-08-459-701-3	Sequence 3, Appl1
3	31.8	6.4	US-08-460-298-3	Sequence 3, Appl1
4	31.8	6.4	US-08-459-174-3	Sequence 3, Appl1
5	31.8	6.4	PCT-US93-06300A-3	Sequence 1, Appl1
6	31.8	6.4	US-08-761-258-1	Sequence 1, Appl1
7	31.8	6.4	US-08-977-306-1	Sequence 1, Appl1
8	31.8	6.4	US-08-343-443B-106	Sequence 106, App
9	30	6.0	US-08-343-443B-1	Sequence 1, Appl1
10	29.6	5.9	US-09-141-000-2	Sequence 2, Appl1
11	29.6	5.9	US-09-103-840A-2	Sequence 2, Appl1
12	29	5.8	US-08-050-132A-8	Sequence 8, Appl1
13	29	5.8	US-08-750-232A-8	Sequence 8, Appl1
14	29	5.8	US-08-815-652A-8	Sequence 8, Appl1
15	29	5.8	US-08-254-353A-8	Sequence 8, Appl1
16	29	5.8	PCT-US92-05374A-8	Sequence 8, Appl1
17	29	5.8	PCT-US95-0708A-8	Sequence 8, Appl1
18	29	5.8	US-08-815-718-1	Sequence 1, Appl1
19	29	5.8	US-08-884-072-4	Sequence 4, Appl1
20	29	5.8	US-09-213-168-4	Sequence 4, Appl1
21	29	5.8	US-08-368-043-1	Sequence 1, Appl1
22	29	5.8	US-08-012-543-1	Sequence 1, Appl1
23	28.8	5.8	PCT-US93-07645A-1	Sequence 1, Appl1
24	28.8	5.8	PCT-US93-07645-1	Sequence 1, Appl1
25	28.8	5.8	US-08-371-121-15	Sequence 15, Appl1
26	28.8	5.8	US-09-127-424-1	Sequence 1, Appl1
27	28.8	5.8	US-09-495-082-1	Sequence 1, Appl1

28	28.8	5.8	3705	5	PCT-US96-03940-7	Sequence 7, Appl1
29	28.8	5.8	5648	5	PCT-US96-03940-1	Sequence 1, Appl1
30	28.6	5.7	838	4	US-09-007-097-1	Sequence 1, Appl1
31	28.6	5.7	838	5	PCT-US94-10457-1	Sequence 1, Appl1
32	28.6	5.7	838	5	PCT-US96-03965-7	Sequence 7, Appl1
33	28.6	5.7	946	2	US-08-816-605-1	Sequence 1, Appl1
34	28.6	5.7	1415	4	US-08-236-918A-7	Sequence 7, Appl1
35	28.6	5.7	1415	4	US-09-150-864A-7	Sequence 142, App
36	28.4	5.7	1652	4	US-08-936-165A-142	Sequence 4, Appl1
37	28.4	5.7	1035	1	US-07-975-526-4	Sequence 5, Appl1
38	28.4	5.7	5438	4	US-08-456-200B-5	Sequence 11, Appl1
39	28.4	5.7	6232	4	US-08-456-200B-11	Sequence 8, Appl1
40	28.4	5.7	7175	2	US-08-455-543A-8	Sequence 8, Appl1
41	28.4	5.7	7175	2	US-08-193-078B-8	Sequence 8, Appl1
42	28.4	5.7	7175	2	US-08-223-305C-8	Sequence 8, Appl1
43	28.4	5.7	7175	2	US-08-149-097B-8	Sequence 8, Appl1
44	28.4	5.7	7175	3	US-08-949-386-8	Sequence 8, Appl1
45	28.4	5.7	7175	3	US-08-450-562-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-287-442-3
; Sequence 3, Application US/08287442
; Patent No. 5670350
; GENERAL INFORMATION:
; APPLICANT: Gaffney, Thomas D.
; APPLICANT: Lam, Stephen T.
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Stein, Jeffrey I.
; APPLICANT: Howell, Charles R.
; APPLICANT: Becker, J. Ole
; TITLE OF INVENTION: Gene Activating Element
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,442
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/087,636
; FILING DATE: 01-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,284
; FILING DATE: 02-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/570,184
; FILING DATE: 08-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELEPHONE/DOCKET NUMBER: CGC 1506/CIP4
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5559 base pairs
; type: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267356
INDIVIDUAL ISOLATE: 5.6 kb EcoRI-HindIII restriction
IMMEDIATE SOURCE: fragment
CLONE: pCIB137
FEATURE:
NAME/KEY: misc-feature
LOCATION: 210..1688
OTHER INFORMATION: /note= "ORF 1, transcribed left to
right"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1906..3633
OTHER INFORMATION: /note= "ORF 2, transcribed left to
right"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4616..4691
OTHER INFORMATION: /note= "glyw, transcribed right to
left"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4731..5318
OTHER INFORMATION: /note= "ORF 3, transcribed right to
left"
US-08-287-442-3

Query Match
Best Local Similarity 50.3%; Score 31.8; DB 1; Length 5559;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 109 GATTGGTGATTTGATGTCGAGAGTGAAGTAAATGCTTATCTGCTGAGGCTG 168
Db 1198 GGTTCCTGGCCTATGTGCGGGGGAATCAACCCGACCTGAGCTGATCCAGCCCG 1257
Qy 169 TGCTGAGATGGCTGTGTGCTGACCTCTCGAGTGCATTTTGCATGTGTAACAGGT 228
Db 1358 AGCTGCGCTTGAATGCGACCAAGCTGCTCCAGGTGCAACAGTCCGGTGTGCGTGA 1317
Qy 229 CTCCTCTGTGGGCAACCAACAAAGAGAGACTTGC 263
Db 1318 CCGCGCGCGGTGCGCATCTGACGGGGAAGAACG 1352

RESULT 2
US-08-459-701-3
Sequence 3, Application US/08459701
Patent No. 5686282
GENERAL INFORMATION:
APPLICANT: Gaffney, Thomas D.
APPLICANT: Lam, Stephen T.
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Stein, Jeffrey I.
APPLICANT: Howell, Charles R.
APPLICANT: Becker, J Ole
TITLE OF INVENTION: Gene Activating Element
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,701
FILING DATE: 02-JUN-1995
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/287,442
FILING DATE: 08-AUG-1994
APPLICATION NUMBER: US 08/087,636
FILING DATE: 01-JUL-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/908,284
FILING DATE: 02-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/570,184
FILING DATE: 08-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5559 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267356
INDIVIDUAL ISOLATE: 5.6 kb EcoRI-HindIII restriction
IMMEDIATE SOURCE: fragment
CLONE: pCIB137
FEATURE:
NAME/KEY: misc-feature
LOCATION: 210..1688
OTHER INFORMATION: /note= "ORF 1, transcribed left to
right"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1906..3633
OTHER INFORMATION: /note= "ORF 2, transcribed left to
right"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4616..4691
OTHER INFORMATION: /note= "glyw, transcribed right to
left"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4731..5318
OTHER INFORMATION: /note= "ORF 3, transcribed right to
left"
US-08-459-701-3

Query Match
Best Local Similarity 50.3%; Score 31.8; DB 1; Length 5559;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 109 GATTGGTGATTTGATGTCGAGAGTGAAGTAAATGCTTATCTGCTGAGGCTG 168
Db 1198 GGTTCCTGGCCTATGTGCGGGGGAATCAACCCGACCTGAGCTGATCCAGCCCG 1257
Qy 169 TGCTGAGATGGCTGTGTGCTGACCTCTCGAGTGCATTTTGCATGTGTAACAGGT 228

Db 1258 AGCTGCGCCCTTGTACTGCGCCACACGCTGTCAGTGCAACAGTCGCTGCGGTGA 1317
OY 229 CTCCTCTCTGTGGGACACACACACAGAGAGTTGC 263
Db 1318 CCGCGCCGCGGTCGATCGACGCGGGAAGAAAGC 1352

RESULT 3
US-08-460-298-3
Sequence 3, Application US/08460298
Patent No. 5686283
GENERAL INFORMATION:
APPLICANT: Gaffney, Thomas D.
APPLICANT: Lam, Stephen T.
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Stein, Jeffrey I.
APPLICANT: Howell, Charles R.
APPLICANT: Becker, J. Ole
TITLE OF INVENTION: Gene Activating Element
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,298
FILING DATE: 02-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/287,442
FILING DATE: 08-AUG-1994
APPLICATION NUMBER: US 08/087,636
FILING DATE: 01-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,284
FILING DATE: 02-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/570,184
FILING DATE: 08-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5559 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEtical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267336
INDIVIDUAL ISOLATE: 5.6 kb EcoRI-HindIII restriction
INDIVIDUAL ISOLATE: fragment
IMMEDIATE SOURCE:
CLONE: pCIB137
FEATURE:

NAME/KEY: misc.feature
LOCATION: 210..1688
OTHER INFORMATION: /note="ORF 1, transcribed left to
OTHER INFORMATION: right"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1906..3633
OTHER INFORMATION: /note="ORF 2, transcribed left to
OTHER INFORMATION: right"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 4616..4691
OTHER INFORMATION: /note="g1yw, transcribed right to
OTHER INFORMATION: left"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 4731..5318
OTHER INFORMATION: /note="ORF 3, transcribed right to
OTHER INFORMATION: left"
US-08-460-298-3

Query Match 6.4%; Score 31.8; DB 1; Length 5559;
Best local similarity 50.3%; Pred. No. 1.3;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

OY 109 GATGTGTGATTTGATGCTGACGGAGAGATGATGAGGCTTATCTGCTGACGCTG 168
Db 1198 GTTCTGCGCTATGTGCGGGGGAATCATCCCGACCTGCTGACCTGATCCACCCG 1257
OY 169 TGCTGAGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 228
Db 1258 AGCTGCGCCCTTGTACTGCGCCACACGCTGTCAGTGCAACAGTCGCTGCGGTGA 1317
OY 229 CTCCTCTCTGTGGGACACACACAGAGAGTTGC 263
Db 1318 CCGCGCCGCGGTCGATCGACGCGGGAAGAAAGC 1352

RESULT 4
US-08-459-174-3
Sequence 3, Application US/08459174
Patent No. 5710031
GENERAL INFORMATION:
APPLICANT: Gaffney, Thomas D.
APPLICANT: Lam, Stephen T.
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Stein, Jeffrey I.
APPLICANT: Howell, Charles R.
APPLICANT: Becker, J. Ole
TITLE OF INVENTION: Gene Activating Element
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,174
FILING DATE: 02-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/287,442
FILING DATE: 08-AUG-1994
APPLICATION NUMBER: US 08/087,636
FILING DATE: 01-JUL-1993

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/908,284
FILING DATE: 02-JUL-1992
PRIORITY APPLICATION DATA: US 07/570,184
APPLICATION NUMBER: 08-AUG-1990
FILING DATE: 08-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP4
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5559 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267356
INDIVIDUAL ISOLATE: 5.6 kb EcoRI-HindIII restriction
IMMEDIATE SOURCE:
CLONE: pCIB137
FEATURE:
NAME/KEY: misc_feature
LOCATION: 210..1688
OTHER INFORMATION: /note="ORF 1, transcribed left to
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1906..3633
OTHER INFORMATION: /note="ORF 2, transcribed left to
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4616..4691
OTHER INFORMATION: /note="glyw, transcribed right to
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4731..5318
OTHER INFORMATION: /note="ORF 3, transcribed right to
US-08-459-174-3
Query Match 6.4%; Score 31.8; DB 1; Length 5559;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 109 GATTGTGATTTGATGTCGACGGAGTGAAGTAAATGAGCTTATCGTCGACAGGCTG 168
DB 1198 GCTTCTCTGCTATGTCGCGGGGGAATATACCCGCAACCTGCTGACCTGATCCAGCCG 1257
QY 169 TCTGAGAGTGCCTGCTGTCGACCCCTCTCTGAGTACATTTGATGTAAGAGGT 228
DB 1258 AACTCGGCTTGAACCTGCGACACCCCTGTTCTCAGGTGCAACAGTCCGGTGTGGGTGA 1317
QY 229 CTCCTCTGGGCAACAAACAAAGAGAGTTC 263
DB 1318 CCGCCGCGGGTGCATCGAGCGGGAAGAAGC 1352
RESULT 5
PCT-US93-06300A-3
; Sequence 3, Application PC/TUS9306300A
; GENERAL INFORMATION:
; APPLICANT: CIBA-GEIGY AG
APPLICANT: Klybeckstrasse 141
APPLICANT: 4002 Basle
APPLICANT: Switzerland
APPLICANT: 125 Tradescant Road
APPLICANT: Chapel Hill, NC 27514
APPLICANT: USA
APPLICANT: 8900 Jeanew Court
APPLICANT: Raleigh, NC 27613
APPLICANT: USA
APPLICANT: Hill, Dwight Steven
APPLICANT: 311 Melanie Lane
APPLICANT: Cary, NC 27511
APPLICANT: USA
APPLICANT: Stein, Jeffrey I.
APPLICANT: 3725 Surry Trail
APPLICANT: Hillsborough, NC 27278
APPLICANT: USA
APPLICANT: Howell, Charles R.
APPLICANT: 805 Avondale
APPLICANT: Bryan, TX 77802
APPLICANT: USA
APPLICANT: Becker, J. Ole
APPLICANT: 6164 Osewego
APPLICANT: Riverside, CA 92506
APPLICANT: USA
APPLICANT: Lyon, James M.
APPLICANT: 120 Marguerite Drive
APPLICANT: Cary, NC 27513
APPLICANT: USA
TITLE OF INVENTION: Gene Activating Element
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06300A
FILING DATE: 02-JUL-1993
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/908,284
FILING DATE: 02-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Spull, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: S-18210/A/CGC1506/PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5559 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267356
INDIVIDUAL ISOLATE: 5.6 kb EcoRI-HindIII restriction
IMMEDIATE SOURCE:
CLONE: pCIB137
```

```

1 TELEFAX: (919) 541-8689
2 INFORMATION FOR SEQ ID NO: 1:
3 SEQUENCE CHARACTERISTICS:
4     LENGTH: 10763 base pairs
5     TYPE: nucleic acid
6     STRANDEDNESS: single
7     TOPOLOGY: linear
8     MOLECULE TYPE: DNA (genomic)
9     HYPOTHETICAL: NO
10    ANTI-SENSE: NO
11    ORIGINAL SOURCE:
12    ORGANISM: Pseudomonas fluorescens
13    STRAIN: CGA267356 (aka MOC6134 and aka BL915)
14    IMMEDIATE SOURCE:
15    CLONE: plasmid pE11
16    FEATURE:
17    NAME/KEY: misc_feature
18    LOCATION: 210..1688
19    OTHER INFORMATION: /product= "methyltransferase"
20    OTHER INFORMATION: /note= "Coding sequence for methyltransferase has homology to
21    OTHER INFORMATION: the cshA and ftrp genes from E. coli and Myxococcus xanthus
22    OTHER INFORMATION: respectively."
23    FEATURE:
24    NAME/KEY: misc_feature
25    LOCATION: 1906..3633
26    OTHER INFORMATION: /product= "sensor kinase"
27    OTHER INFORMATION: /note= "Coding sequence for sensor kinase has homology to
28    OTHER INFORMATION: rcsC, frzE, and bvgS genes of E. coli, M. xanthus, and
29    OTHER INFORMATION: Bordetella pertussis, respectively."
30    FEATURE:
31    NAME/KEY: misc_RNA
32    LOCATION: complement (4616..4691)
33    OTHER INFORMATION: /product= "tRNA"
34    OTHER INFORMATION: /note= "(complementary DNA strand) Homology to glyW from E. coli"
35    OTHER INFORMATION:
36    FEATURE:
37    NAME/KEY: misc_feature
38    LOCATION: complement (4731..5318)
39    OTHER INFORMATION: /product= "response
40    OTHER INFORMATION: "CDP-diacylglycerol-3-phosphate-3-phosph
41    OTHER INFORMATION: atidylyltrans."
42    OTHER INFORMATION: /note= "Coding sequence for
43    OTHER INFORMATION: CDP-diacylglycerol-3-phosphate-3-phosphatidylyltran
44    OTHER INFORMATION: se has homology to pgsA."
45    FEATURE:
46    NAME/KEY: misc_feature
47    LOCATION: complement (5574..7397)
48    OTHER INFORMATION: /product= "UVR exonuclease subunit
49    OTHER INFORMATION: C"
50    OTHER INFORMATION: /note= "Coding sequence for UVR exonuclease subunit C has
51    OTHER INFORMATION: homology to uvrC."
52    FEATURE:
53    NAME/KEY: misc_feature
54    LOCATION: complement (7400..8041)
55    OTHER INFORMATION: /function= "response
56    OTHER INFORMATION: regulator/transcription activator"
57    OTHER INFORMATION: /product= "gacA (aka gafA)"
58    OTHER INFORMATION: /note= "Coding sequence for gacA (aka gafA) has homology to
59    OTHER INFORMATION: uvvY and gacA genes of E. coli and Ps. fluorescens,
60    OTHER INFORMATION: respectively."
61    US-08-761-256-1
62    Query Match 6.4%; Score 31.8; DB 1; Length 10763;
63    Best Local Similarity 50.3%; Pred. No. 1.9;
64    Matches 78; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
65    Oy 109 GATGGGTGATTTGATGCTGACGGGAGTGAAGATAATGGCTTATTCCTGCGACAGGCTG 168
66    Db 1198 GGGTCTCTGCGCTATGTCGGGGGGAATDACCAGCACTGCTGACCTGATTCAGACGCG 1257
67    Oy 169 TCGTAGAGTGGCTGGTCTGACCACTCTCTCGATGACATTTTGCATGTGTADAGGCT 228
68    Db 1258 AGCTGGCGCTTGAACTCGACCACTGCTTCCTGAGGTGCAACAGCTCCGGTGTGGGGTGA 1317

```

QY 229 CTCCTCTGGGGCAGACACACAGAGAGTTGC 263
DB 1318 CCGCCGCCGGGTGCATCGACCGGAGAAAGAGC 1352

RESULT 7
US-08-977-306-1
Sequence 1, Application US/08977306
Patent No. 595348
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Gaffney, Thomas D.
APPLICANT: Yorkewitz, Nancy
APPLICANT: Stafford, Jill M.
TITLE OF INVENTION: Genetically Modified Pseudomonas Strains
TITLE OF INVENTION: With Enhanced Biocontrol Activity
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 595348artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,306
FILING DATE:
CLASSIFICATION: A25
AUTHOR/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 36,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10763 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267356 (aka MOCG134 and aka BL915)
IMMEDIATE SOURCE:
CLONE: Plasmid pE11
FEATURE:
NAME/KEY: misc-feature
LOCATION: 210..1688
OTHER INFORMATION: /product= "methyltransferase"
OTHER INFORMATION: /note= "Coding sequence for methyltransferase has homology to
OTHER INFORMATION: the cheR and frzF genes from E. coli and Myxococcus xanthus,
OTHER INFORMATION: respectively."
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1906..3633
OTHER INFORMATION: /product= "sensor kinase"
OTHER INFORMATION: /note= "Coding sequence for sensor kinase has homology to the
OTHER INFORMATION: rscC, frzE, and frzF genes of E. coli, M. xanthus, and
OTHER INFORMATION: Bordetella pertussis, respectively."
NAME/KEY: misc-RNA
LOCATION: complement (4616..4691)
OTHER INFORMATION: /product= "tRNA"

OTHER INFORMATION: /note= "(complementary DNA strand) Homology to glyw from E
OTHER INFORMATION: coli."
FEATURE:
NAME/KEY: misc-feature
LOCATION: complement (4731..5318)
OTHER INFORMATION: /product= "CDP-diacylglycerol-3-phosphate-3-phosph
OTHER INFORMATION: atidylyltrans."
OTHER INFORMATION: /note= "Coding sequence for
OTHER INFORMATION: CDP-diacylglycerol-3-phosphate-3-phosphatidylyltran
OTHER INFORMATION: se has homology to pgsa."
FEATURE:
NAME/KEY: misc-feature
LOCATION: complement (5574..7397)
OTHER INFORMATION: /product= "uvr exonuclease subunit
OTHER INFORMATION: C"
OTHER INFORMATION: /note= "Coding sequence for uvr exonuclease subunit C has
OTHER INFORMATION: homology to uvrC."
FEATURE:
NAME/KEY: misc-feature
LOCATION: complement (7400..8041)
OTHER INFORMATION: /function= "response
OTHER INFORMATION: regulator/transcription activator"
OTHER INFORMATION: /product= "gacA (aka gafa)"
OTHER INFORMATION: /note= "Coding sequence for gacA (aka gafa) has homology to
OTHER INFORMATION: uvrI and gacA genes of E. coli and Ps. fluorescens,
OTHER INFORMATION: respectively."

US-08-977-306-1
Query Match 6.4%; Score 31.8; DB 2; Length 10763;
Best Local Similarity 50.3%; Pred. No. 1.9; 77; Indels 0; Gaps 0;
Matches 78; Conservative 0; Mismatches 0

QY 109 GATTGTGATTTGATGCTGACGGAGTGAAGTAAAGCCCTATCTGCTGACAGGCTG 168
DB 1198 GGTTCCTGCGCCCTATGCTGCGGGGAAATCACCCGCAACCTGACCTGATCCAGCCCG 1257
QY 169 TGTGAGGATGCGCTGCTGCTGACCCCTCTCGAGTAGCATTTTTCATGTATACAGGGT 228
DB 1258 AGCGCGCCCTTGAACTGCGCACACAGCTGTTCCAGGTGCAACAGTGGTGTGGGGTGA 1317
QY 229 CTCCTCTGGGGCAGACACACAGAGAGTTGC 263
DB 1318 CCGCCGCCGGGTGCATCGACCGGAGAAAGAGC 1352

RESULT 8
US-08-343-443B-106/c
Sequence 106, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmazes, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploogastel, Beatrice
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
TITLE OF INVENTION: TRANSLOCATIONS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:

RESULT 9
 US-08-343-443B-1/c
 Sequence 1, Application US/08343443B
 Patent No. 5968734
 GENERAL INFORMATION:
 APPLICANT: Aurias, Alain
 APPLICANT: Delattre, Olivier
 APPLICANT: Desmaze, Chantal
 APPLICANT: Melot, Thomas
 APPLICANT: Peter, Martine
 APPLICANT: Ploougastel, Beatrice
 APPLICANT: Thomas, Gilles
 APPLICANT: Zucman, Jessica
 TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
 TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
 TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF
 TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
 NUMBER OF SEQUENCES: 129
 TRANSLATIONS
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Welser & Associates
 STREET: 230 South Fifteenth Street
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk

```

RESULT 10
US-09-141-000-2/c
Sequence 2, Application US/09141000
Patent No. 6054295
GENERAL INFORMATION:
APPLICANT: Chem. Fang
TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR
FILE REFERENCE: 19999Y
CURRENT APPLICATION NUMBER: US/09/141,000
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 500
TYPE: PRT
ORGANISM: Human
US-09-141-000-2

Query Match
Best Local Similarity 84%, Pred. No. 1.9:
Matches 38; Conservative 122; Mismatches 284; Indels 0; Gaps 0

OY      44 TTTTGGAACATTTGTCTGACGCACAAGATAATATATATGCAACACTCGCCTTATTTTA 103
       |::: ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      456 TYD..S.B...YW...SHYYTMT..K...MDK..BMC..MBSR.D..BYMTTA.Y... 397
OY      104 AATTGATGGTGTGATTTGATTCCTACGCGGAGTGACAGTAAAGTAAAGCCCTTCTCTCCA 163
       |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 396 . . . MCRYW . . . DY . S . RH . T . D . H . M . BT . H . KRSHSN . T . TM . AB . . M . 337

Qy 164 GCGTGTGTGAGGATGCGCTGTGTGCGACCCCTCCGAGTAGGATTTTCATGTGTAAAC 223

Db 336 BM . MMYBY . . . TYR . . . CT . YSD . HK . RH . TPB . MHHR . . SYNB . C KWS . 277

Qy 224 AAGGTCCTCCCTCGGGGAGCAACACAAAGAAAGATTTCTTACGACAAAGACAGGCGC 283

Db 276 . . . SK . HC . S . SS . C . DMWC . BB . YH . HC . AA . TM . HC . KC . . . KTR . MH . TB . 217

Qy 284 GGAATGATCTCTCCATCTTGGAACAGCCCTGGTTATCCCAATGCTCGAAGAGGCGTA 343

Db 216 . RSB . MA . MMT . S . . . S . R . SS . SH . YIMRA . YCTBYH . YBC . M . KCBM . GAK . YGT 157

Qy 344 TGGCAGCGCTCCCAAGAGCTCTGCAGCGTCGACCTTGGGGGTGACAGCTCTGTCTTTGCC 403

Db 156 . GSWIYYUA . G . MAT . GSR . NC ATM . . TMGT . GSTBCHDKCST . KCGB . G . YM . T 97

Qy 404 TCGCTGATTAAGGCGCGCTGAAGCCAGCCCACTGCTGCCCAATATCCACAGCGCATTTGGG 463

Db 96 N . KR . GM . TSK . K . CSTHG . NMSTG . MCC . CTCT . K . . SH . CMTDAC . T . C . CDSHTSSC 37

Qy 464 GGTTTTCCATGAGCGCACCTTCGCGGAGGCAAG 497

Db 36 S . SDDS . TH . CSS . SS AM . SCGSSC . DYHNS 3

RESULT 11
US-09-103-840A-2/C
Sequence 2, Application US/09103840A

```

? GENERAL INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: FRASER, Claire M.
? APPLICANT: VENER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? TITLE OF INVENTION: TUBERCULOSIS
? PUBLICATION NUMBER: 36677_00
? CURRENT APPLICATION NUMBER: 52/09/103,840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO: 2
? LENGTH: 4403765
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? FEATURE:
? OTHER INFORMATION: CDC 1551
? OTHER INFORMATION: "n" bases at various positions throughout the sequence
? OTHER INFORMATION: represent a, t, c or g
? US-09-103-840A-2

```

[illegible]

RESULT 12
US-08-050-132A-8/C
; Sequence 8, Application US/08050132A

Patient No. 5661007
 GENERAL INFORMATION:
 APPLICANT: Wozney, John M.
 APPLICANT: Celisny, Anthony
 TITLE OF INVENTION: BAP-9 COMPOSITIONS
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: Legal Affairs - 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: MA
 COUNTRY: US
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/050,132A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kapinos, Ellen J.
 REGISTRATION NUMBER: 32,245
 REFERENCE/DOCKET NUMBER: GI 5186A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 470 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL LINE: W138 (genomic DNA)
IMMEDIATE SOURCE:
LIBRARY: human genomic library
CLONE: lambda 111-1
POSITION IN GENOME:
UNITS: bp
FEATURE:
NAME/KEY: exon
LOCATION: 1..470
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 124..453
FEATURE:
NAME/KEY: mRNA
LOCATION: 1..470
IS-08-050-132A-8

Query Match	5.8%	Score 29	DB 1	Length 470	
Best Local Similarity	52.0%	Pred. No. 2.9			
Matches	65	Conservative	0	Mismatches	60
				Indels	0
				Gaps	0
QY	332	AGAGAGGCGCTAAGCGCCAGTCCCTCCCAAGCTCTGCGACGTCGACCTTTGGGGGTGGGACACGC	391		
Db	347	ACACACAGCAGGCGCTTGTCCACCTTTGTGGGGAACTTGMATGACACAGGGTGTGCACATA	288		
QY	392	TGCTGCTTGTCTGCGCTGATPAAGCGCCGTGAAGCCACGACCCACTGTGTCGCCAATAATACC	451		
Db	287	GGGTATTTCGTGGCGCTGCACATGTCACACCAAGGGGAAGAAGAGAGCGCCCTTACACTCG	228		
XY	452	CAGCC	456		

Fri Oct 11 09:30:38 2002

us-09-489-101a-12_copy_1_500.rni

Page 10

LOCATION: 1..456
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 124..453
FEATURE:
NAME/KEY: mRNA
LOCATION: 1..470
US-08-815-652B-8

Query Match 5.8%; Score 29; DB 3; Length 470;
Best Local Similarity 52.0%; Pred. No. 2.9;
Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 332 AGAGAGTGTATGCGAGTCTCTCCAGAGCTCTGCGAGCTCTGGGGGTGACAGTC 391
DB 347 ACACAGAGGAGCTTGGCCACCTTTGTGGGAACTTGAGATGACACAGGCTCTGACGATA 288
QY 392 TCGTGTCTGCTGCGGTGATTAAGGCGCGTGAAGCCAGCCACTGCTGCCAAATTCACC 451
DB 287 GCGGTCTGCTGCGGTGATTAAGGCGCGTGAAGGGAAGAGCAGCGCTTACACTCG 228
QY 452 CAGCC 456
DB 227 TAGGC 223

RESULT 15
US-08-254-353A-8/C
Sequence 8, Application US/08254353A
Patent No. 6287816
GENERAL INFORMATION:
APPLICANT: Rosen, Vicki A.
APPLICANT: Mooney, John M.
APPLICANT: Celeste, Anthony J.
APPLICANT: Song, Jeffrey
TITLE OF INVENTION: BMP-9 COMPOSITIONS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: Legal Affairs - 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: US
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,353A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,245
REFERENCE/DOCKET NUMBER: GI 5186B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 470 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL LINE: W138 (genomic DNA)

IMMEDIATE SOURCE:
LIBRARY: human genomic library
CLONE: Lambda 111-1
POSITION IN GENOME:
UNITS: bp
FEATURE:
NAME/KEY: exon
LOCATION: 1..470
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
NAME/KEY: mat_peptide
LOCATION: 124..453
FEATURE:
NAME/KEY: mRNA
LOCATION: 1..470
US-08-254-353A-8

Query Match 5.8%; Score 29; DB 4; Length 470;
Best Local Similarity 52.0%; Pred. No. 2.9;
Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 332 AGAGAGTGTATGCGAGTCTCTCCAGAGCTCTGCGAGCTCTGGGGGTGACAGTC 391
DB 347 ACACAGAGGAGCTTGGCCACCTTTGTGGGAACTTGAGATGACACAGGCTCTGACGATA 288
QY 392 TCGTGTCTGCTGCGGTGATTAAGGCGCGTGAAGCCAGCCACTGCTGCCAAATTCACC 451
DB 287 GCGGTCTGCTGCGGTGATTAAGGCGCGTGAAGGGAAGAGCAGCGCTTACACTCG 228
QY 452 CAGCC 456
DB 227 TAGGC 223

Search completed: October 10, 2002, 19:26:08
Job time : 2506.73 secs

RESULT	1
LOCUS	A0587521
DEFINITION	A0587521 675 bp DNA linear GSS 07-JUN-1999
ACCESSION	CITRBI-EI-2643G14.TF CITRBI-EI Homo sapiens genomic clone 2643G14,
VERSION	A0587521
KEYWORDS	A0587521.1 GI:5014201
SOURCE	GSS.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (base 1 to 675)
AUTHORS	Zho,S., Adams,M.D., Nierman,W., Malek,J., Shiyaya,H., Simon,M. and Venet,J.C.
TITLE	Use of BAC End Sequences from Caltech Libraries for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other_GSSs: CITRBI-EI-2643G14.TR Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hpe@ligr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.ligr.org/vdb/hungten/bac_end_search/bac_end_search.html . Seq primer: M13-21 Class: BAC ends.
FEATURES	location/Qualifiers
source	1..675 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="2643G14" /clone_lib="CITRBI-EI" /sex="male" /cell_type="sperm" /vector="pBeloBAC11, Site_1: EcoRI, Site_2: EcoRI;

Caltech Human BAC Library D"
BASE COUNT 173 a 134 c 162 g 206 t
ORIGIN

Query Match 48.8%; Score 243.8; DB 12; Length 675;
Best Local Similarity 99.2%; Pred. No. 4.5e-65;
Matches 245; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACCTGTCGACCTATTTTGGACTATGCTTCGATACAGCTTTATGAGACATTGTC 60
DB 429 AACCTGTCGACCTATTTTGGACTATGCTTCGATACAGCTTTATGAGACATTGTC 488
QY 61 AGGCAAAAGTATATATATGCAAACTACAGCTTTATTTTAAATAGATTGCTGAT 120
DB 489 AGGCAAAAGTATATATATGCAAACTACAGCTTTATTTTAAATAGATTGCTGAT 548
QY 121 TTGATGTCGACGGAGTATGAGTATGAGCTTATGCTTCGACAGGCTGCTGAGGATG 180
DB 549 TTGATGTCGACGGAGTATGAGTATGAGCTTATGCTTCGACAGGCTGCTGAGGATG 608
QY 181 CTTGCTGTCGACCTCTCCGAGTATGCTTCGATGCTGACAGGCTCTCCCTGTCG 240
DB 609 CTTGCTGTCGACCTCTCCGAGTATGCTTCGATGCTGACAGGCTCTCCCTGTCG 668
QY 241 GCACAAAC 247
DB 669 GCACAAAC 675

RESULT 2 723 bp DNA linear GSS 29-SEP-2000
LOCUS AZ344779
DEFINITION 1M0079E03F Mouse 10kb plasmid UNGCM library Mus musculus genomic
clone UNGCM0079E03 F, DNA sequence.
ACCESSION AZ344779
VERSION AZ344779.1 GI:10424016
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 723) Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Dunn, D., Aoyagi, A., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, R.
and Wright, D., Weis, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah
Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0079 row: E column: 03
Seq primer: CGTGTAAACGACGCGCACT
Class: plasmid ends
High quality sequence stop: 723.
Location/Qualifiers
1. 723

FEATURES
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UNGCM0079E03"
/clone_lib="Mouse 10kb plasmid UNGCM library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g11473211419b/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 163 a 172 c 206 g 182 t
ORIGIN

Query Match 8.4%; Score 42.2; DB 12; Length 723;
Best Local Similarity 63.1%; Pred. No. 0.052; 38; Indels 0; Gaps 0;
Matches 65; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 2 AACCTGTCGACCTATTTTGGACTATGCTTCGATACAGCTTTATGAGACATTGTC 61
DB 714 AACCTGTCGACCTATTTTGGACTATGCTTCGATACAGCTTTATGAGACATTGTC 655
QY 62 GCACAAAGTATATATATGCAAACTACAGCTTTATTTTAA 104
DB 654 GCACAAAGTATATATATGCAAACTACAGCTTTATTTTAA 612

RESULT 3 931 bp DNA linear GSS 02-JUN-2000
LOCUS AZ128097
DEFINITION OSJNB0086H01r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
clone OSJNB0086H01r, DNA sequence.
ACCESSION AZ128097
VERSION AZ128097.1 GI:8204877
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartioideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 931) Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: GGAACAGCTATGACGANG
Class: BAC ends
High quality sequence start: 19
High quality sequence stop: 615.
Location/Qualifiers
1. 931

FEATURES
source
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNB0086H01r"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACind190; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of

RESULT	8			
BEST74740		339 bp	mRNA	linear
LOCUS	BEST74740			EST-20-SEP-2000
DEFINITION	MRA-U00021-30660-008-e09 U00021 Homo sapiens cDNA, mRNA sequence.			
ACCESSION	BEST74740			
VERSION	BEST74740.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			

REFERENCE
1 (bases 1 to 781)
AUTHORS
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duvál, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
unpublished (2000)

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Alcalá, J., Vrehalov, J., White, R., Visions, T., Karamycheva, S. A., Tsai, J., Bougari, O., Kirkness, E., Utecherak, T., Van Aken, S., Rönning, C. M., Fraser, C. M., Martin, G. B., Tanksey, S. D. and Giovannoni, J.	Generation of ESTs from tomato fruit tissue, breaker stage (2002)	Unpublished (2002)	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html This clone is available through the Clemson University Genomics Institute

```

FEATURES
source
Location/Qualifiers
1. .790
/organism="lycopersicon esculentum"
/cultivar="PA496"
/db_xref="taxon:4081"
/clone="G1EG61K23"
/clone_lib="tomato breaker fruit"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/notes="Vector: plasmidsKmc2adap; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
Sequencing: The institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit).
Fruits were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT
205 a 151 c 168 g 266 t
ORIGIN

```

Query Match	Score	DB 10	Length	790
Best Local Similarity	49.4%	Pred.	10	
Matches 118	Conservative	0	Mismatches 120	Indels 1
			Gaps	1
QY	37	TACACCTTTATGGCAACATTTGACAGGCAAAAGATATATATGCAAACTGTACGCTTT	96	
Db	657	TACACTTTGTGACATGACATCCGACGCAAAAGAACATGATATAGCCATTTTATACCTT	598	
QY	97	TATTTTAAATAGATGTGCTGTTTAAATGCTGACGGAGTGAAGATATGCGCTTATTC	156	
Db	597	TATCATTAATATGATTTGTGATACATTTCTGTGTACATCATTTCTTGTGACGCTTTACAC	538	
QY	157	TGCTGACAGCTGTGCTGAGAGATGAGGCTGCTGTGCGACCCCTGCTGATATTTTGGCA	215	
Db	537	AACGATCAACCTTTTCTGTAATTCATCCCTTTTTCCTCCATTCCTCTAGACCTTCCA	478	
QY	216	TGTGTAAACAGGCTTCCTCTGCGGCAACACAAACAAAGAAATTTGCTAAGGACAA	274	
Db	477	CGTAAACAGGAGGATCATTCATTAAGACTTTGAACACATGATATGATGCTCAAAATTTATA	419	

RESULT 12	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AL034894		376 bp mRNA linear EST 29-DEC-1999	AL034894	AL034894	house mouse. Mus musculus	house mouse. Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	1 (bases 1 to 376)	Harrison,S.M., Dunwoodie,S.L., Arkell,R.M., Lechach,H. and Beadlington,R.S.	Isolation of novel tissue-specific genes from cDNA libraries
AL034894		clone 528_9h17 5', mRNA sequence.	AL034894.1	GI:6646520	EST.					

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 84.933 seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-12_COPY_1_500

Perfect score: 500
Sequence: 1 aagcttggtgcacatctattt.....ccctgcgcgagccaaagaag 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_032802:*
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001a.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001b.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	8372	22	AA011119	Human small cell 1
2	33.6	6.7	2811	22	AA020201	Corynebacterium gl
3	33.6	6.7	2811	22	AA020201	Nucleotide sequenc
4	33.4	6.7	2811	22	AA020201	Human gene express
5	33.3	6.6	1214	11	AA008551	Rat BRL-3A binding
6	32.8	6.6	373	22	AA008825	Human CDNA clone (
7	32.6	6.6	1952	22	AA015237	Human CDNA sequenc
8	32.6	6.5	1366	21	AA016216	Human prostate can
9	32.6	6.5	32248	22	AA020412	Human nervous syst

C	10	32.6	6.5	32248	22	AA037122	Human musculoskele
C	11	32.6	6.5	32248	22	AA026795	Human genomic DNA
C	12	32.4	6.5	1671	21	AA079977	Human secreted pro
C	13	32.4	6.5	5003	22	AA039887	Genomic sequence #
C	14	32.4	6.5	5003	22	AA039888	Genomic sequence #
C	15	32.4	6.5	5003	22	AA090258	Human digestive sy
C	16	32.4	6.5	5003	22	AA090259	Human digestive sy
C	17	32.2	6.4	36267	23	AB019424	Drosophila melanog
C	18	32.2	6.4	480	21	AA028235	Human colon cancer
C	19	31.8	6.4	321	21	AA028235	Human colon cancer
C	20	31.8	6.4	5559	15	AA055260	Human ORF2220
C	21	31.8	6.4	5559	15	AA055260	Restriction fragme
C	22	31.8	6.4	5559	18	AA092122	5.6 Kb E11 Pseudom
C	23	31.8	6.4	5559	18	AA092122	Pseudomonas fluores
C	24	31.8	6.4	5559	18	AA092122	E11 genomic DNA fr
C	25	31.8	6.4	5559	19	AA092122	5.6 Kb E11-Hind
C	26	31.8	6.4	5559	19	AA092122	Human immune/haema
C	27	31.8	6.4	5559	19	AA092122	Pseudomonas fluores
C	28	31.6	6.3	10763	20	AA092122	P. fluorescens 11k
C	29	31.6	6.3	6016	23	AB02857	Drosophila melanog
C	30	31.6	6.3	8807	23	AB02857	Drosophila melanog
C	31	31.6	6.3	17137	24	AB032190	Human immune syste
C	32	31.4	6.3	15362	22	AA036491	Human musculoskele
C	33	31.4	6.3	15362	22	AA036491	Human musculoskele
C	34	31.2	6.2	259	21	AA011925	Human secreted pro
C	35	31.2	6.2	893	21	AA060591	PEMT nucleotide se
C	36	31.2	6.2	1238	22	AA022424	Rat phosphatidylet
C	37	31.2	6.2	1238	22	AA022424	Human bone marrow
C	38	31.2	6.2	10689	22	AA067355	C glutamicum codin
C	39	31.2	6.2	10708	22	AA065569	Human immune/haema
C	40	31.2	6.2	349980	22	AA065531	Human immune/haema
C	41	31.2	6.2	251	19	AA011496	C glutamicum codin
C	42	31.2	6.2	251	19	AA011496	Human biallelic po
C	43	31.2	6.2	251	19	AA011496	Human biallelic po
C	44	31.2	6.2	251	19	AA011496	Human biallelic po
C	45	31.2	6.2	251	19	AA011496	Human biallelic po

ALIGNMENTS

RESULT 1	AA011119	standard; DNA: 8372 BP.
XX	AA011119	
AC	AA011119	
XX	AA011119	
DT	24-SEP-2001	(first entry)
XX	24-SEP-2001	
DE	Human small cell lung cancer associated gene, SOX21.	
XX	Human: small cell lung cancer; therapy: hCAP: nucleic acid: NA:	
KW	melanoma; cancer: colon; breast; head; neck; transitional cancer;	
KW	leiomyosarcoma; synovial sarcoma; cytostatic; SOX21; ds.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	1167..1997
FT	FT	/*tag= a
FT	FT	/product= "Human SOX21 protein"
PN	WO20015349-A2.	
PD	26-JUL-2001.	
XX	19-JAN-2001; 2001WO-US02015.	
PT	21-JAN-2000; 2000US-0489101.	
XX	(LUDW-) LUDWIG INST CANCER RES.	
XX	(SLOK) SLOAN KETTERING INST CANCER RES.	
PA	(CORR) CORNELL RES FOUND INC.	

XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
 XX WPI; 2001-457597/49.
 DR P-PSDB; AAE05814.
 XX
 PT Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e.g. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 PS Claim 57; Page 100-105; 152pp; English.

XX The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterized by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterized by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterized by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is a small cell lung cancer associated gene
 CC designated as NT-SCLC-10, encoding human SOX21 protein.
 XX

SQ Sequence 8372 BP; 1996 A; 2143 C; 2104 G; 2124 T; 5 other;

Query Match 100.0%; Score 500; DB 22; Length 8372;
 Best Local Similarity 100.0%; Pred. No. 1.1e-152;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

YY 1 AAGCTTGGTCCATCTATTTGGCATATGCTGATACAGCTTATGGGAACATTGTG 60
 DB 1 AAGCTTGGTCCATCTATTTGGCATATGCTGATACAGCTTATGGGAACATTGTG 60
 YY 61 AGCAAAAGATATATATATGCAAACTCTACGCTTTATTTAAATTTAGATGGTGTAT 120
 DB 61 AGCAAAAGATATATATATGCAAACTCTACGCTTTATTTAAATTTAGATGGTGTAT 120
 YY 121 TTGATGCTGACGGAGTGAAGATATGCTTATCTGCTGACGCTGCTGATGATG 180
 DB 121 TTGATGCTGACGGAGTGAAGATATGCTTATCTGCTGACGCTGCTGATGATG 180
 YY 181 CCGTGTGTCACCTCCCTGAGTATGATTTGATGATGATGATGATGATGATGATG 240
 DB 181 CCGTGTGTCACCTCCCTGAGTATGATTTGATGATGATGATGATGATGATGATG 240
 YY 181 CCGTGTGTCACCTCCCTGAGTATGATTTGATGATGATGATGATGATGATGATG 240
 DB 181 CCGTGTGTCACCTCCCTGAGTATGATTTGATGATGATGATGATGATGATGATG 240
 YY 241 GCACACAAACAAAGAGATGTTGTAAGCAAGACAGAGTCCGGAATGATCTCCAT 300
 DB 241 GCACACAAACAAAGAGATGTTGTAAGCAAGACAGAGTCCGGAATGATCTCCAT 300
 YY 301 TGGAAACAGCCCTGGGCTTACTCCATGCTGAGAGAGTGTATGCGCATCTCCAGA 360
 DB 301 TGGAAACAGCCCTGGGCTTACTCCATGCTGAGAGAGTGTATGCGCATCTCCAGA 360
 YY 361 GCTGTGAGTGTGATGTTGGGGGGTGGACAGTCTGCTGCTGCTGATGATGATG 420
 DB 361 GCTGTGAGTGTGATGTTGGGGGGTGGACAGTCTGCTGCTGCTGATGATGATG 420
 YY 421 GAAAGCCAGCAACTGCTGCGCAAAATACCAAGCCGATTTGGGGTTCCTCCAG 480
 DB 421 GAAAGCCAGCAACTGCTGCGCAAAATACCAAGCCGATTTGGGGTTCCTCCAG 480
 YY 481 CCTGCGCGGAGCCCAAGAAG 500
 DB 481 CCTGCGCGGAGCCCAAGAAG 500

RESULT 2
 AAD20201/C
 ID AAD20201 standard; DNA; 2811 BP.
 XX

AC AAD20201;
 XX
 DT 03-JAN-2002 (first entry)
 XX
 DE Corynebacterium glutamicum glucose 6-phosphate isomerase (pgi) DNA.
 XX
 KW L-amino acid: fermentation; coryneform bacteria; animal nutrition;
 KW glucose 6-phosphate dehydrogenase; zwf; human medicine;
 KW foodstuff industry; glucose 6-phosphate isomerase; pgi;
 KW pharmaceutical industry; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 FH Key Location/Qualifiers
 FT CDS 373..2025
 FT /tag= a
 FT /product= "Glucose 6-phosphate isomerase (pgi)"

MO20017095-A1.

27-SEP-2001.

05-JUL-2000; 2000WO-EP06303.

20-MAR-2000; 2000US-0531269.

(DEGS) DEGUSA AG.
 (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 (UYNA) UNIV NAT IRELAND.

Burke K, Salm H, Eggeling L, Moritz B, Dunican LK, McCormack A;
 Stapleton C, Moeckel B, Thierbach G;
 WPI: 2001-639130/73.
 P-PSDB; AAE12461.

PT Preparing L-amino acids by fermenting coryneform bacteria transformed
 PT with the glucose 6-phosphate dehydrogenase gene is particularly useful
 PT to produce L-lysine and L-threonine -
 XX
 PS Example 4; Page 49-52; 63pp; English.

XX The present invention relates to a method for preparing L-amino acids by
 CC fermenting coryneform bacteria, comprising fermenting the L-amino acid
 CC producing bacteria in which at least the glucose 6-phosphate
 CC dehydrogenase (zwf) gene is amplified, and concentrating and isolating
 CC the L-amino acid produced. The L-amino acids produced are used in animal
 CC nutrition, human medicine, foodstuffs industry and the pharmaceuticals
 CC industry. The present sequence is Corynebacterium glutamicum strain
 CC AS019 glucose 6-phosphate isomerase (pgi) DNA.
 XX

SQ Sequence 2811 BP; 592 A; 744 C; 764 G; 711 T; 0 other;

Query Match 6.7%; Score 33.6; DB 22; Length 2811;
 Best Local Similarity 50.0%; Pred. No. 3.3;
 Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

YY 180 GCGTGTGCGCACCTCTGAGTATGATTTGCAATGATGATGATGATGATGATG 239
 DB 180 GCGTGTGCGCACCTCTGAGTATGATTTGCAATGATGATGATGATGATGATG 239
 YY 2804 GCGCGCTAGAGTACTCTGAGAGCTTTTGTATTTTGTGATATCTACCTATG 2745
 DB 2804 GCGCGCTAGAGTACTCTGAGAGCTTTTGTATTTTGTGATATCTACCTATG 2745
 YY 240 GGCACACAAACAAAGAGATGTTGTAAGACAGAGAGTCCGGAATGATCTCCA 299
 DB 240 GGCACACAAACAAAGAGATGTTGTAAGACAGAGAGTCCGGAATGATCTCCA 299
 YY 2744 GGCATAAACCCATGAGATATAGTGCAGAAACGAAATGAAACCAACACAC 2685
 DB 2744 GGCATAAACCCATGAGATATAGTGCAGAAACGAAATGAAACCAACACAC 2685
 YY 300 TTGGAACAGCCCTGGGCTTACTCCATGCTGAGAGAGTGTATGCG 347
 DB 300 TTGGAACAGCCCTGGGCTTACTCCATGCTGAGAGAGTGTATGCG 347
 YY 2684 CCCACACCAACCGAAGACTTCCACACAGAGAGAAACCTTCGACGCG 2637
 DB 2684 CCCACACCAACCGAAGACTTCCACACAGAGAGAAACCTTCGACGCG 2637

RESULT 3
 AAF84712/C
 ID AAF84712 standard; DNA; 2811 BP.
 XX

AAE84712;
29-JUN-2001 (first entry)
Nucleotide sequence of a bacterial glucose-6-phosphate isomerase.
Corynebacterium bacteria; glucose-6-phosphate isomerase; gpi; metabolic flux; pentose phosphate cycle; L-amino acid; nucleotide production;
vitamin production; L-amino acid production; ss.
Corynebacterium glutamicum.

Key Location/Qualifiers
CDS 373..2025 /tag=a
FT /product= "glucose-6-phosphate isomerase"
EP1087015-A2.
28-MAR-2001.
23-AUG-2000; 2000EP-0118052.
15-SEP-1999; 99US-0396478.
(DEGS) DEGUSSA-HUELS AG.
(UYNA-) UNIV NAT IRELAND.
Dunican LR, McCormack A, Stapelton C, Burke K, O'Donohue M;
Marx A, Moockel B;
WPI: 2001-267593/28.
P-RSDB: AAEB8033.

Novel nucleotide sequence from coryneform bacteria encoding
glucose-6-phosphate isomerase, attenuation of which increases metabolic
flux through pentose phosphate cycle in the bacteria, producing L-amino
acids

Claim 4: Page 13-16; 33pp; English.

The present sequence encodes a coryneform bacterial glucose-6-phosphate
isomerase (gpi). After attenuation of the gpi gene, coryneform bacteria
exhibit an improved metabolic flux through the pentose phosphate cycle
and produce nucleotides, vitamins, and in particular L-amino acids,
particularly L-leucine and L-tryptophan, in an improved manner.
Attenuation may be achieved by reducing or switching off either the
expression of the gpi gene or the catalytic properties of the enzyme
protein. Microorganisms, in particular Corynebacterium glutamicum in
which the gpi gene is attenuated or switched off, optionally in
combination with amplification of further genes, are useful for
producing nucleotides, vitamins or L-amino acids.

Sequence 2811 BP; 592 A; 744 G; 764 G; 711 T; 0 other;

Query Match 6.7%; Score 33.6; DB 22; Length 2811;
Best Local Similarity 50.0%; Pred. NO. 3.3;
Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0.

180 GCCTGGTCTGCACCCCTCGAGTAGATTTTGCAATGTGTAACAGGCTGCCCTGG 239
| | | | | | | | | | | | | | | | | |
Db 2804 GGCGGCTGTGAAGTACTCTCGAGAAGCTTTTGAATTCTTGTGATCATCTACATCTGC 2745
| | | | | | | | | | | | | | | | | |
QY 240 GGACACAACAANGAAGATGTCMAAGAACAAAGCAGSTGGGAAATGANTCCCCA 299
| | | | | | | | | | | | | | | | | |
Db 2744 GGCTATAAACCTATGAGATATACCTCGCAACGAAAGTAACCACCAATCATCACCGAGC 2685
| | | | | | | | | | | | | | | | | |
QY 300 TTGGACACAGCCTGGGCTTACTCATGTGGCTGAGAGAGCTGTATGGC 347
| | | | | | | | | | | | | | | | | |
Db 2684 CCCACACCAACGAAAGCTTCACACAGGAAGAAAGAACTTCTGACGGC 2637
| | | | | | | | | | | | | | | | | |

XX	RESULT 4
XX	AAZ15378
XX	AAZ15378 standard; cDNA; 774 BP.
AC	
XX	AAZ15378;
DT	
XX	12-OCT-1999 (first entry)
DE	
XX	Human gene expression product cDNA sequence SEQ ID NO:2847.
KW	Human; gene; gene expression product; diagnosis; therapy; probe;
KM	detection; mapping; tissue typing; profiling; forensic; cancer;
XX	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
OS	Homo sapiens.
PN	
XX	M0938972-AZ.
PD	
XX	05-AUG-1999.
PF	
XX	26-JAN-1999; 99MO-US01619.
PR	
XX	03-APR-1998; 98US-0080666.
PR	28-JAN-1998; 98US-0072910.
PR	24-FEB-1998; 98US-0075954.
PR	31-MAR-1998; 98US-0080114.
PR	03-APR-1998; 98US-0080515.
XX	
PA	(CHIR) CHIRON CORP.
XX	(HYSE-) HYSEQ INC.
P1	Crtvenjakov R, Dickson M, Drmanac R, Drmanac S;
P1	Escobedo J, Garcia PD, Garcia V, Giese K, Imis MA;
P1	Joscoe WL, Kassam A, Kennedy GC, Kita D, Labat J;
P1	Lemson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
P1	Stache-Crain B, Suduth-Klinger J, Williams LJ.
DR	WPI; 1999-494092/41.
XX	
PT	Novel human genes and their expression products which are
PT	differentially expressed in different cell types
PS	
XX	Claim 1; Page 1377; 2479p; English.
XX	
CC	The present invention describes a library of human polynucleotides
CC	comprising the sequences given in AAZ12532 to AAZ1779. Also described is
CC	a method of detecting differentially expressed genes correlated with the
CC	cancerous state of a mammalian cell, comprising detecting at least one
CC	differentially expressed gene product in a test sample from a cell
CC	suspected of being cancerous, where the gene product is encoded by one
CC	of the 5348 polynucleotide sequences given in AAZ12532 to AAZ1779. The
CC	polynucleotides can be used as a source of primers and probes, which can
CC	be used for a variety of purposes, e.g., detection of expression levels,
CC	tissue typing or profiling, forensics, genetic analysis and
CC	detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC	can be used for raising antibodies for experimental, diagnostic and
CC	therapeutic purposes. The polynucleotides may also be used to construct
CC	arrays for diagnostics (which may be used to determine function of an
CC	encoded protein); and to detect differences in expression levels between
CC	two cells (e.g., to identify abnormal or diseased tissue in a human, to
CC	identify a genetic predisposition or susceptibility to a disease such as
CC	cancer). The polynucleotides or the invention are especially used in the
CC	diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC	and lung cancer. The polynucleotides can also be used to screen for
CC	peptide analogues and antagonists.
XX	
SQ	Sequence 774 BP; 233 A; 119 C; 123 G; 251 T; 48 other;
Query Match	6.7%; Score 33.4; DB 20; Length 774;
Best Local Similarity	47.8%; Ped. No. 1.9;
Matches 97; Conservative	0; Mismatches 106; Indels 0; Gaps 0;

CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAF57303 represent sequences used in the exemplification of the present
CC invention.

CC Sequence 1366 BP; 262 A; 387 C; 327 G; 386 T; 4 other;

Query Match Best Local Similarity 59.6%; Score 32.6; DB 21; Length 1366;

Matches 53; Conservative 1; Mismatches 35; Indels 0; Gaps 0;

OY 42 CTTATGGACATTTGTGACGCAAAAGTATATATGCAACTGACGCTTTATTT 101
DB 275 CTTCTCTTATTTTCTTATGATATGATTTTATTTTATTTCTTTCTTTT 334
OY 102 TAAATGATTTGGTGTGATTTGATGCTGA 130
DB 335 TGGTTTGATTTGCTTTGTTGAGGAGA 363

RESULT 9

ABR20412
ID ABR20412 standard; DNA; 32248 BP.

XX ABR20412;

XX 23-JAN-2002 (first entry)

XX Human nervous system related polynucleotide SEQ ID NO 12743.

XX Human: nocotropic; neuroprotective; cytostatic; dermatological; vitruclide;
XX immunosuppressive; antiinflammatory; anti-HIV; antihypertensive; vulnery;
XX antiparinsonian; antisticking; antianemic; antileukemic; cancer;
XX antileukemic; hepatoprotective; cerebroprotective; antileukemic;
XX antileukemic; antidiabetic; antileukemic; antileukemic; antileukemic;
XX antileukemic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAR-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217486.
PR 26-JUL-2000; 2000US-0218299.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 14-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0242617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.

Qy	43	TTTTGTGGCAACATTTGTGAGCAAAAGTATATATATGCAAACTCTAGCCCTTTTATTTT	102
Db	25793	TCTAAAGGCTCCGCTGCATGCAAAAAGTATGATTCATTAATGATCAATTTGTATAGCTTTTCTCT	25851
Oy	103	AAATTAGATTGGTGTGATT	121
Db	25852	CAATCTGCTTTTGTCACT	25870
RESULT 10			
AA137122/C			
ID	AA137122	standard; DNA; 32248 BP.	
AC	AA137122:		
DT	08-JAN-2002	(first entry)	
XX			
XX			
XX			
De		Human musculoskeletal system related polynucleotide SEQ ID NO 3487.	
KM		Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KM		antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;	
KM		vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KM		cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KM		neurological disease; infection; human; secreted protein;	
KM		musculoskeletal system; ds.	
XX			
OS		Homo sapiens.	
XX			
PN		WO20015367-A1.	
PD			
XX			
XX			
XX			
PE	17-JAN-2001:	2001WO-US01338.	
XX			
PR	31-JAN-2000:	2000US-0179065.	
PR	04-FEB-2000:	2000US-0180628.	
PR	24-FEB-2000:	2000US-0184664.	
PR	02-MAR-2000:	2000US-0186350.	
PR	16-MAR-2000:	2000US-0189874.	
PR	17-MAR-2000:	2000US-0190076.	
PR	18-APR-2000:	2000US-0198133.	
PR	19-MAY-2000:	2000US-0205515.	
PR	07-JUN-2000:	2000US-0209467.	
PR	28-JUN-2000:	2000US-0214886.	
PR	30-JUN-2000:	2000US-0215135.	
PR	07-JUL-2000:	2000US-0216647.	
PR	11-JUL-2000:	2000US-0217487.	
PR	14-JUL-2000:	2000US-0217496.	
PR	26-JUL-2000:	2000US-0218290.	
PR	14-AUG-2000:	2000US-0220963.	
PR	14-AUG-2000:	2000US-0224518.	
PR	14-AUG-2000:	2000US-0224519.	
PR	14-AUG-2000:	2000US-0225213.	
PR	14-AUG-2000:	2000US-0225214.	
PR	14-AUG-2000:	2000US-0225266.	
PR	14-AUG-2000:	2000US-0225267.	
PR	14-AUG-2000:	2000US-0225268.	
PR	14-AUG-2000:	2000US-0225447.	
PR	14-AUG-2000:	2000US-0225757.	
PR	14-AUG-2000:	2000US-0225758.	
PR	14-AUG-2000:	2000US-0225759.	
PR	16-AUG-2000:	2000US-0226429.	
PR	22-AUG-2000:	2000US-0226561.	
PR	22-AUG-2000:	2000US-0226566.	
PR	23-AUG-2000:	2000US-0227182.	
PR	30-AUG-2000:	2000US-0228094.	
PR	30-AUG-2000:	2000US-0228094.	

Db 6397 CAATCTGCCTTGTGTCAGT 6379

RESULT 11

AA526795

ID AA526795 standard; DNA: 32248 BP.

XX

AC AA526795;

XX

DT 07-NOV-2001 (first entry)

XX

DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1769.

XX

KW Human; immunosuppressive; antiarthritic; ds; antirheumatic;

KW cytoskeletal; cardiant; vasotropic; cerebroprotective; nootropic;

KW neuroprotective; antibacterial; virucide; fungicide; optalmatological;

KW vulnerrary; secreted protein; rheumatoid arthritis;

KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;

KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;

KW corneal infection; wound healing; epithelial cell proliferation;

KW skin ageing; food additive; preservative; antiproliferative.

XX

OS Homo sapiens.

XX

PN WO20015322-A2.

XX

PD 02-AUG-2001.

XX

XX 17-JAN-2001; 2001WO-US01341.

XX

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184564.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0186874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-APR-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225256.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226661.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 14-SEP-2000; 2000US-0233066.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234597.

PR 25-SEP-2000; 2000US-0234598.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235634.

PR 27-SEP-2000; 2000US-0235636.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241121.

PR 20-OCT-2000; 2000US-0241185.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246533.

PR 08-NOV-2000; 2000US-0246534.

PR 08-NOV-2000; 2000US-0246535.

PR 08-NOV-2000; 2000US-0246536.

PR 08-NOV-2000; 2000US-0246537.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.

17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250191.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256119.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259578.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-488783/53.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Disclosure; SEQ ID NO 1769; 980bp; English.

The invention relates to isolated nucleic acid molecules and proteins encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunosays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence is a genomic DNA encoding a partial novel secreted protein of the invention.

Query Match 6.5%; Score 32.6; DB 22; Length 32248;
 Best Local Similarity 63.3%; Pred. No. 27;
 Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

0Y 43 TTTATGGAACTTTGTCAGGCAAGTAAATTAATGCAAACTCTAGGCTTTATTT 102
 DB 25792 TCTAAAGGCTCTGTCATGAAAGATATGATGATCAATTTGATGCTTTCTCT 25851
 0Y 103 AATTAAGTTGTTGATTT 121
 DB 25852 CAATCTGCTTTGTCAGT 25870

RESULT 12
 AAC79977/c
 ID AAC79977 standard; cDNA; 1671 BP.
 XX
 AC AAC79977;
 XX
 DT 12-FEB-2001 (first entry)
 XX
 DE Human secreted protein encoding cDNA for gene 30.
 XX
 KW Secreted protein; human; immunosuppressive; antiarthritic; antihemetic;
 KW antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; vitruide; fungicide;
 KW ophthalmological; vulnereary; gene therapy; treatment; autoimmune disease;
 KW hyperproliferative disorder; cardiovascular disorder; ocular disorder;
 KW cerebrovascular disorder; nervous system disorder; infection; skin aging;
 KW wound healing; epithelial cell proliferation; transplantation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058357-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 23-MAR-2000; 2000MO-US07723.
 XX
 PR 26-MAR-1999; 99US-0126506.
 PR 07-JAN-2000; 2000US-0174852.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI; 2000-611704/58.
 DR P-PSDB; AAB45054.
 XX
 PT Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers -
 XX
 PS Claim 1a; Page 357; 418bp; English.

This invention describes novel isolated nucleic acid molecules (I) encoding a human secreted protein (II) which have immunosuppressive, antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiac, vasotropic, cerebroprotective, neurotropic, neuroprotective, antibacterial, virucide, fungicide, ophthalmological and vulnereary activity and can be used for gene therapy. (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to (II) can also be used in alleviating symptoms associated with the disorders and in diagnostic immunosays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities.

Query Match 6.5%; Score 32.4; DB 21; Length 1671;
 Best Local Similarity 56.6%; Pred. No. 6.2;
 Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Sequence 1671 BP; 481 A; 294 C; 325 G; 571 T; 0 other;

```
OY 17 ATTTGGACTANGCTTCGATACACCTTTATGGAACATTTGTGACGAACATATATA 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1198 ATGTTAGACTATTCAGTACGTATATATCTTGTACTTGTAGAGGAGATNACCA 1139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 77 ATGGAACACTACGCGCTTTATTTAAATATGATTTGGTGATTT 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1138 ACAGGACACGACTACTCTTAACCTGCTTGGTTTGAGTACT 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
ID AAS39887 standard; DNA, 5003 BP.
AC AAS39887;
XX 17-DEC-2001 (first entry)
XX
DE Genomic sequence #306 encoding human colon associated polypeptide.
XX
KM Human; colon cancer; congenital abnormality; infection; colitis;
KM Inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KM Intestinal inflammatory disorder; malabsorption syndrome; gastritis;
KM sigmoid disease; antibacterial; antiviral; antiinflammatory;
KM cytostatic; ds.
XX
OS Homo sapiens.
XX
PN W0200155302-A2.
XX
FD 02-AUG-2001.
XX
FE 17-JAN-2001; 2001WO-US01240.
XX
FA 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198113.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0230963.
PR 26-JUL-2000; 2000US-0230964.
PR 14-AUG-2000; 2000US-0234518.
PR 14-AUG-2000; 2000US-0234519.
PR 14-AUG-2000; 2000US-0235213.
PR 14-AUG-2000; 2000US-0235256.
PR 14-AUG-2000; 2000US-0235257.
PR 14-AUG-2000; 2000US-0235258.
PR 14-AUG-2000; 2000US-0235270.
PR 14-AUG-2000; 2000US-0235271.
PR 14-AUG-2000; 2000US-0235275.
PR 14-AUG-2000; 2000US-0235279.
PR 18-AUG-2000; 2000US-0236681.
PR 22-AUG-2000; 2000US-0236681.
PR 22-AUG-2000; 2000US-0237182.
PR 23-AUG-2000; 2000US-0237009.
PR 30-AUG-2000; 2000US-0238924.
PR 01-SEP-2000; 2000US-0239287.
PR 01-SEP-2000; 2000US-0239343.
PR 01-SEP-2000; 2000US-0239344.
PR 01-SEP-2000; 2000US-0239345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234297.
PR 25-SEP-2000; 2000US-0234398.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0235637.
PR 29-SEP-2000; 2000US-0235638.
PR 29-SEP-2000; 2000US-0235639.
PR 29-SEP-2000; 2000US-0236362.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240560.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
```


PT polynucleotides encoding digestive system antigens, useful for
 PT diagnosing, treating, preventing and/or prognosing disorders of the
 PT digestive system, particularly cancer and cancer metastases -

PS disclosure: SEQ ID NO 3834; 986bp; English.

XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human digestive system antigens. These can be used in the
 CC diagnosis, treatment and prevention of digestive system disorders,
 CC including cancer, Meckel's diverticulum, bacterial or parasitic
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
 CC ulcerative colitis. The present sequence is a genomic DNA fragment
 CC encoding a digestive system antigen of the invention.

XX
 SO Sequence 5003 BP: 1809 A; 870 C; 1047 G; 1277 T; 0 other:

Query Match 6.5%; Score 32.4; DB 22; Length 5003;
 Best Local Similarity 60.0%; Pred. No. 11;

Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 18 TTTTGACTATGCTTGCATACAGCTTTATGAGACATTGTCAGGCAAAAGTATATTA 77

Db 4916 TTTGGCCCTATTTCGACATACATTGGAACCTTTCCAGTACTAAGATAATA 4857

QY 78 TGGCAACTCTACGCTTTTATTTAATT 107

Db 4856 TTGGAGCTTTATGCAATGTATTAAATT 4827

Search completed: October 10, 2002, 15:17:13
 Job time : 141 secs

28.4	5.7	7218	1	US-08-232-653-14	Sequence 14, Appl
28.2	5.6	1379	2	US-08-437-607A-1	Sequence 1, Appl
28.2	5.6	1380	2	US-08-437-607A-1	Sequence 4, Appl
28.2	5.6	2344	3	US-08-883-852A-2	Sequence 2, Appl
28.2	5.6	4402765	4	US-09-103-840A-2	Sequence 2, Appl
28	5.6	443	1	US-08-470-179-109	Sequence 109
28	5.6	1357	6	5504194	Patent No. 5504194
28	5.6	17001	4	US-09-816-088-3	Sequence 3, Appl
27.8	5.6	12001	1	US-08-458-568A-11	Sequence 11, Appl
27.6	5.5	1060	2	US-08-385-935A-14	Sequence 14, Appl
27.6	5.5	1656	3	US-08-659-168-17	Sequence 17, Appl
27.6	5.5	1656	3	US-08-655-427-17	Sequence 17, Appl
27.6	5.5	1656	3	US-08-655-427-17	Sequence 17, Appl
27.6	5.5	1656	4	US-09-338-326-17	Sequence 17, Appl
27.6	5.5	2241	3	US-09-144-759-19	Sequence 19, Appl
27.6	5.5	2241	3	US-09-144-759-19	Sequence 19, Appl
27.6	5.5	2301	3	US-09-144-759-21	Sequence 21, Appl
27.6	5.5	2676	3	US-09-212-971-11	Sequence 11, Appl

FEATURE:
NAME/KEY: CDS
LOCATION: 195, 1310
SEQUENCE DESCRIPTION: SEQ ID NO: 3
US-09-043-937A-3

Query Match
Best Local Similarity 45.2%; Score 32.2; DB 4; Length 1556;
Matches 118; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 238 GGGGCGGGGGGACGAAAGGCTCCCGAATTTTTCACGCGGACGCGG 297
DB 478 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 537
QY 298 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
DB 538 AGCAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
QY 358 GTTCGAGGGGGGACCAAGCCGCGGCGGCGGCGGCGGCGGCGGCGG 417
DB 598 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 657
QY 418 AGCTAGAGCCGACGAGCTGTGATGCGACCTGTCGAGAGAACTCATGCGGAGA 477
DB 658 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 717
QY 478 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 498
DB 718 AAACGAGAACTGTGTCTCT 738

RESULT 2
US-08-937-580-20/c
Sequence 20, Application US/08937580
Patent No. 6013510
GENERAL INFORMATION:
APPLICANT: Harris, James M.
TITLE OF INVENTION: Identification of a DNA Region
TITLE OF INVENTION: Potentially Useful for the Detection of Mycobacterium
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Bection Dickinson and Company
STREET: 1 Bection Drive
CITY: Franklin Lakes
STATE: New Jersey
COUNTRY: USA
ZIP: 07417-6800
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,580
FILING DATE: 25-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3690/5510-13
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-848-9228
TELEFAX: 201-848-9228
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORIGINAL SOURCE:

ORGANISM: Mycobacterium gastr
US-08-937-580-20

Query Match
Best Local Similarity 53.2%; Score 31.6; DB 3; Length 311;
Matches 67; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 203 CCTAAGTACTTACACGCTCAGTACGCGGCGGCGGCGGCGGCGGCGG 262
DB 148 CCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 89
QY 263 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 322
DB 88 ACCGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 29
QY 323 CAGCAG 328
DB 28 CAGCAG 23

RESULT 3
US-09-336-039-20/c
Sequence 20, Application US/09336039
Patent No. 6291176
GENERAL INFORMATION:
APPLICANT: Harris, James M.
TITLE OF INVENTION: Identification of a DNA Region
TITLE OF INVENTION: Potentially Useful for the Detection of Mycobacterium
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Bection Dickinson and Company
STREET: 1 Bection Drive
CITY: Franklin Lakes
STATE: New Jersey
COUNTRY: USA
ZIP: 07417-6800
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/336,039
FILING DATE: 18-Jun-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/937,580
FILING DATE: 25-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3690/5510-13
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-848-9228
TELEFAX: 201-848-9228
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Mycobacterium gastr
US-09-336-039-20

Query Match
Best Local Similarity 53.2%; Score 31.6; DB 4; Length 311;
Matches 67; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 203 CCTAAGTACTTACACGCTCAGTACGCGGCGGCGGCGGCGGCGGCGG 262

Db 148 CCGTAGGCAATTCGACGCGGATGCGATCAGAGCAGCTGGACACGACGCGCTT 89
QY 263 CCCGAACCTTTTTCAGCCAGCCGACGAGGCGCTCGTAATGATGGCCAGGCGC 322
11
Db 88 ACCCGAGTTCGTCGCCCGCACGACCGCCGCGTGCATGCGTACGATGCGTGGCGCG 29
QY 323 CATCAC 328
11111
Db 28 CATCGC 23

RESULT 4
US-08-937-560-19/C

; Sequence 19, Application US/08937580
; Patent No. 6013510
; GENERAL INFORMATION:
; APPLICANT: Harris, James M.
; TITLE OF INVENTION: Identification of a DNA Region
; TITLE OF INVENTION: Potentially Useful for the Detection of Mycobacterium
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Becton Dickinson and Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07417-6800
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 25-SEP-1997
; APPLICATION NUMBER: US/08/937,580
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3690/5510-13
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-847-7166
; TELEFAX: 201-848-9228
; INFORMATION FOR SEQ ID NO: 19:
; LENGTH: 311 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium gastr1 and Mycobacterium
; US-08-937-560-19

Query Match 6.0%; Score 30; DB 3; Length 311;
Best Local Similarity 47.6%; Pred. No. 1.2;

Matches 60; Conservative 8; Mismatches 58; Indels 0; Gaps 0;

QY 203 CCTAAGTAACCTTACACGCTCAGTCAAGCAAGAGGCGCTGGGTGAACGAAGGCT 262
11
Db 148 CCGTAGGCAATTCGACGCGGATGCGATCAGAGCAGCCGAGACGACGAGCGCTT 89
QY 263 CCCGAACCTTTTTCAGCCAGCCGACGAGGCGCTCGTAATGATGGCCAGGCGC 322
11
Db 88 ACCCGAGTTCGTCGCCCGCACGACCGCCGCGTGCATGCGTACGATGCGTGGCGCG 29
QY 323 CATCAC 328
11111
Db 28 CATCGC 23

RESULT 5
US-09-336-039-19/C
; Sequence 19, Application US/09336039
; Patent No. 6291176
; GENERAL INFORMATION:
; APPLICANT: Harris, James M.
; TITLE OF INVENTION: Identification of a DNA Region
; TITLE OF INVENTION: Potentially Useful for the Detection of Mycobacterium
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Becton Dickinson and Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07417-6800
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 18-Jun-1999
; APPLICATION NUMBER: US/09/336,039
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,580
; FILING DATE: 25-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3690/5510-13
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-847-7166
; TELEFAX: 201-848-9228
; INFORMATION FOR SEQ ID NO: 19:
; LENGTH: 311 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium gastr1 and Mycobacterium
; US-09-336-039-19

Query Match 6.0%; Score 30; DB 4; Length 311;
Best Local Similarity 47.8%; Pred. No. 1.2;

Matches 60; Conservative 8; Mismatches 58; Indels 0; Gaps 0;

QY 203 CCTAAGTAACCTTACACGCTCAGTCAAGCAAGAGGCGCTGGGTGAACGAAGGCT 262
11
Db 148 CCGTAGGCAATTCGACGCGGATGCGATCAGAGCAGCCGAGACGACGAGCGCTT 89
QY 263 CCCGAACCTTTTTCAGCCAGCCGACGAGGCGCTCGTAATGATGGCCAGGCGC 322
11
Db 88 ACCCGAGTTCGTCGCCCGCACGACCGCCGCGTGCATGCGTACGATGCGTGGCGCG 29
QY 323 CATCAC 328
11111
Db 28 CATCGC 23

RESULT 6
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:

```

? APPLICANT: FLEISCHMAN, Robert D
? APPLICANT: WHITE, Owen R.
? APPLICANT: FRASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? FILE REFERENCE: 24366-20007.00
? CURRENT APPLICATION NUMBER: US/09/103,840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ. ID NOS.: 2
? SOFTWARE: PatentIn Ver. 2.1
? SEQ. ID NO. 1
? LENGTH: 4411529
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? OTHER INFORMATION: H37Rv
? US-09-103-840A-1

```

Query Match:	6.0%;	Score 30;	DB 4;	Length 441529;
Best Local Similarity:	52.4%;	Pred. No. 47;		
Matches	66;	Conservative	0;	Mismatches 60;
			Indels	0;
			Gaps	0;
QY	316	CAGGGCGCATACATCGCCAAACGTGTAATACAGGGCTCCCTGGGTTCCGAGGGGGGAGCA	375	
Db	2151568	CAGAGACCCAGCCCGCGTAATCGTCTCATACATCCGCCCTCCGCAAGCCGTGTTGTAATGC	2151627	
QY	376	ACACCCCAACCCCGGGGAATCCGAGCAGGTAATTAAGGGGCCCTAAGAGCCCAAGGACA	435	
Db	2151628	GGCGACACCGCTTGCAGGATCCGCTGGTTATCTCCGGCTCTGGGGTGGGGTCCCACTGC	2151687	
QY	436	CTGTGTGA	441	
Db	2151688	CAGGGA	2151693	

RESULT 7
 US-08-379-580-3
 : Sequence 3, Application US/08379580
 : Patent No. 6180362
 :
 GENERAL INFORMATION:
 APPLICANT: Duchesne, Marc
 APPLICANT: Schweighoffer, Fabien
 APPLICANT: Tocque, Bruno
 TITLE OF INVENTION: Peptides Inhibiting Ras Protein
 TITLE OF INVENTION: Activity, Preparation and Use Thereof
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Rhone-Poulenc Rorer Inc.
 STREET: 500 Arcola Rd. 3c43
 CITY: Collegeville
 STATE: PA
 COUNTRY: USA
 ZIP: 19002
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/379,580
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR93/00772
 FILING DATE: 28-Jul-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/09433
 FILING DATE: 03-Jul-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: SMITH, JULIE K.
 REGISTRATION NUMBER: 38,619
 REFERENCE/DOCKET NUMBER: ST92049-US
 TELECOMMUNICATION INFORMATION:

? TELEPHONE: (610)454-3839
 ? TELEFAX: (610)454-3808
 ? INFORMATION FOR SEQ ID NO: 3:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 1548 base pairs
 ? TYPE: nucleic acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: cDNA
 ? ORIGINAL SOURCE:
 ? ORGANISM: Homo sapiens
 ? FEATURE:
 ? NAME/KEY: CDS
 ? LOCATION: 1..1548
 ? US-08-379-580-3

```

0Y      68 CCCCTCCTCCCCCTCCGCCAATGCTGGCGTGGAGGAGGAGGAGGAGGT 122
       111111111111111111111111111111111111111111111111111
Db      14 CCCCCTCCTCCCCCTCCGCCCTACTGCTGCCTTTGGGGGCGGACCTCGGACACT 68

Query Match: 5.9%; Score 29.4; DB 4; Length 1548;
Best Local Similarity 70.9%; Pred. No. 4.1;
Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

RESULT 8
Sequence 7, Application US/08190687B
Patent No. 5760203
GENERAL INFORMATION:
APPLICANT: Wong, Gall I.
APPLICANT: Martin, George
APPLICANT: McCormick, Francis P.
APPLICANT: Rubinfeld, Bonnie
APPLICANT: O'Rourke, Edward C.
APPLICANT: Clark, Robin
TITLE OF INVENTION: GAP Gene Sequences
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,687B
FILING DATE: 02-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/774,644
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/260,807
FILING DATE: 21-OCT-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/230,761
FILING DATE: 10-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 27527/31898
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEX: 312/474-0448
TELEFAX: 25-3856
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

```

LENGTH: 4307 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 119..3259
US-08-190-687B-7

Query Match 5.9%; Score 29.4; DB 1; Length 4307;
Best Local Similarity 70.9%; Pred. No. 6.6;
Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 68 CCCCTGCCCCCCCCCAATCTGCTGCGGGGGTGGGGGGGT 122
Db 522 CCCCTGCCCCCCCCCTTACCTGCCCCCTTGGGGGGCGGACAGT 576

RESULT 9

US-08-276-852-156/c
Sequence 156, Application US/08276852
Patent No. 5652138
GENERAL INFORMATION:
APPLICANT: Burton, Dennis R
APPLICANT: Barbas, Carlos F
APPLICANT: Lerner, Richard A
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,852
FILING DATE: 18-JUL-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCRI452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ. ID NO.: 156:
SEQUENCE CHARACTERISTICS:
LENGTH: 13254 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-276-852-156

Query Match 5.9%; Score 29.4; DB 1; Length 13254;
Best Local Similarity 50.3%; Pred. No. 11;
Matches 72; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 116 GGGGGTACCTCTCAGGTTTCCTTGAACCTTTGAAACCTTAATGGGCT 175
Db 8968 GGGCAGCAGTACCTCAGCATTTTCTCTGTAATCTTGACCCAGCATCAAGCAGGCGC 8909
Qy 176 CTGAGTGGGCTCTGGAGCTCCCGCTCTTAATCTTTACCACTGACCTAGGCCAA 235
Db 8908 GGTAGTACCTCCAGCATATCTCTGCTGAGCTTGTGCGCCACACACAGTAAAT 8849
Qy 236 GAGGGGCTGGGGTGAACGAAG 258
Db 8848 ACGGACCTTGGGGCCAGGAAG 8826

RESULT 10

US-08-276-852-170
Sequence 170, Application US/08276852
Patent No. 5652138
GENERAL INFORMATION:
APPLICANT: Burton, Dennis R
APPLICANT: Barbas, Carlos F
APPLICANT: Lerner, Richard A
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,852
FILING DATE: 18-JUL-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCRI452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ. ID NO.: 170:
SEQUENCE CHARACTERISTICS:
LENGTH: 13254 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-276-852-170

Query Match 5.9%; Score 29.4; DB 1; Length 13254;
Best Local Similarity 50.3%; Pred. No. 11;
Matches 72; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 116 GGGGGTACCTCTCAGGTTTCCTTGAACCTTTGAAACCTTAATGGGCT 175
Db 4287 GGGCAGCAGTACCTCAGCATTTTCTCTGTAATCTTGACCCAGCATCAAGCAGGCGC 4346

QY	176	CGTAGGCGGCGCTCGGAGACCTCCGCAAGTAACCTTTACACAGCTCACTAGGCCAAA	235
Db	4347	GGTAGTAGACCTCCACGATATTCCTGCGATAGGCTTTGTCTGCGCCACACACAGATAAT	4400
QY	236	GAGGGGCGCGGGGTGACGAAGAAG	258
Db	4407	ACGGACCTTGGGGCCGAGGAAG	4429

RESULT 11
US-08-899-575-156/c

Sequence 156, Application US/088995/5
Patent No. 5770440

GENERAL INFORMATION:
APPLICANT: Burton, Dennis R
APPLICANT: Barbas, Carlos F
APPLICANT: Lerner, Richard A
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170

CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESS: Patent Counsel
STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,
CITY: Malibu, CA 90265
STATE: CA
CITY: La Jolla

COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,575
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/276,852
FILING DATE: 18-JUL-1994
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992

ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCH1452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-6312
TELEFAX: 619-554-2937

INFORMATION FOR SEQ ID NO: 156:
SEQUENCE CHARACTERISTICS:
LENGTH: 13254 base pairs
type: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
IS-08-899-575-156

Query Match	5.9%	Score 29.4	DB 1	length 13254
Best Local Similarity	50.3%	Pred. No. 11		
Matches	72	Conservative	0	Mismatches 71
			Indels	0
			Gaps	0

QY	116	GGGGGGCTACCTCTCAGGTTGCTGCTTTTAACTTTTGGAAACCCATTAATGGTGCCCT	175
DB	8968	GGGCGAGGATACCTCAGCATTTGTTCTGTATCTTGAACCCAGATACAGCAGCGCC	8909
QY	176	CTGATGGGCTCTGTGGACTCCGCGCTCTTAAGTAACCTTTCACAGTCACTAGGCCAA	235
DB	8908	GGTATGTGCCCTCCACGATATCCCTGCGCATAGGCGCTTGTTGTGGCGCCACACAGTAAT	8849

QY	236	GAGGGCGTGGGGTGAACGAAG	258
Db	8848	ACGGACCTTTGGGGCCAGGAAG	8826

RESULT 12
US-08-899-575-170

Sequence 1/0, Application US/088995/3
Patent No. 5770440

GENERAL INFORMATION:
APPLICANT: Burton, Dennis R
APPLICANT: Barbas, Carlos F
APPLICANT: Lehner, Richard A
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170

CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, office of
ADDRESS: Patent Counsel
STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,
STREET: Mail Drop TPC8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,575
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/276,852
FILING DATE: 18-JUL-1994
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992

ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCH1452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 170:
SEQUENCE CHARACTERISTICS:
LENGTH: 13254 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
US-08-899-575-170

Query Match	Similarity	Score	DB 1	Length
Best Local	50.3%	Pred. No. 11;		13254;
Matches	72;	Conservative	0;	Mismatches 71; Indels 0; Gaps 0;
QY	116	GGGCGGTCACCTCCTCAGTTCGTCTTTCAAACTTTTGAACCCATAATGGTGCT	175	
Db	4287	GGCGAGGCGATACCTCAGCATTTGTTCTGTGTAATCTTGACCCAGAGATAACAAGAGGCGC	4246	
QY	176	CGAATGSGGCTCTGTGACATCCCGGCTCTTACAGTAACCTTTACACAGTCATAGGCCAA	235	
Db	4347	GGTAATGTAGCCCTCCAGCATATCCCTCGCATATAGGCTTTGTCTGTGGCCCAACACACAGTAAT	4406	
QY	236	GAGGGGCGCTGGGGTGAACGAAG	258	

us-09-489-101a-11_copy_1_500.rni

```

1  GENERAL INFORMATION:
2  APPLICANT:
3  TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
4  TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
5  NUMBER OF SEQUENCES: 170
6  COMPUTER READABLE FORM:
7  MEDIUM TYPE: floppy disk
8  COMPUTER: IBM PC compatible
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EBO)
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: PCT/US95/08743
13 FILING DATE: 11-JUL-1995
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: US 08/276,852
16 FILING DATE: 18-JUL-1994
17 INFORMATION FOR SEQ ID NO: 156:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 13254 base pairs
20 TYPE: nucleic acid
21 STRANDEDNESS: double
22 TOPOLOGY: circular
23 MOLECULE TYPE: DNA (genomic)
24
25 PCT-US95-08743-156

```

	Query Match	Best Local Similarity	5.9%; Score 29.4; DB 5; Length 13254;
	Matches 72; Conservative	0; Mismatches 71; Indels 0; Gaps 0;	
Qy	116	GGGGGGGACCTCCTCAGGGTTGGTCTTTCAAAATTTTGAACCCGTAATTTGSGTCC	175
Db	8968	GGCGAGGCAATGACCTCAGCATTTGTTCCTGTATTTTACCCGACGATACAGAGGCGC	8909
Qy	176	CTGAGTGGGGCTGTGTGACTCCCGCTCCCTAATTACTTACAGAGCTATAGGCCAA	235
Db	8908	GCTGTGATGACCTCAGCATATCTCCGCAATAGGCTTTGTCTCGGCCACACACAGTAAT	8849
Qy	236	GAGGGGCGTGGGTGAACGAAAG	258
Db	8848	ACGGACCTTGGGGCCGAGGAAAG	8826

Search completed: October 10, 2002, 18:44:40
Job time : 1790.73 secs

Contact: Yoshhide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
url:http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normal full-length cDNA sublibraries for rapid discovery of new
genes. *Genome Res.* 10 (10), 1617-1630 (2000)
waga, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Tomoda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and
Hayashizaki, Y.

genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon.location/Qualifiers>

```
/organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone="167C22"
/clone_lib="G"
```

BASE COUNT

```

/note="Genoscope sequence ID : C0AG167BB1ILP1-end : T7"
299 a 169 c 146 g 259 t 127 others

```

14 a	19 c	19 g	9 t
------	------	------	-----

Query Match	8.28;	Score 41;	DB 9;	Length 61;
Best Local Similarity	82.58;	Pred. No. 0.79;		
Matches 47;	Conservative	0;	Mismatches 10;	Indels 0;
			Gaps	0;

QY 444 GGCACCTGTTCCAGAGAACAATCATCAGGTGGCGAAGCCCCGGGGTTCCTGTAATT 500
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GCACCACTCGACGGAACGCATCAGGTGAGAGAAGCCCGCGAATCCCGCCACTT 57

RESULT 6			
CNS020HV/c			
LOCUS	1000 bp	DNA	linear
DEFINITION	Tetradon nigroviridis genome survey sequence T7 end of clone		

ACCESSION	AL214492
VERSION	AL214492.1
KEYWORDS	GI:7873311
SOURCE	GSS: genome survey sequence
ORGANISM	Tetraodon nigrivittatus.
	Tetraodon nigrivittatus

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

t-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C.,

TITLE	CHARACTERIZATION AND REPEAT ANALYSIS OF THE COMPACT GENOME OF THE FRESHWATER PUFFERFISH <i>TETRAODON NIGROVIRIDIS</i>
ABSTRACT	...
INTRODUCTION	...
MATERIALS AND METHODS	...
RESULTS	...
DISCUSSION	...
CONCLUSION	...
REFERENCES	...
ACKNOWLEDGMENTS	...
APPENDIX	...
INDEX	...

Published
cases 1 to 1000)

TITLE	Saurth, W. and Weissenbach, J. Human gene number estimate provided by genome wide analysis using method of Saurth et al. 1994
-------	---

JOURNAL REFERENCE	Unpublished 3 (bases 1 to 1000,

TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a larger scale of direct and indirect sequencing of the <i>hcr</i> gene.

10

genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES	location/qualifiers
source	1. . 1000

```

/organism="Tetraodon nigroviridis
/db_xref="taxon:99883"

```

```
/clone="167C22"  
/clone_1lb="G"
```

BASE COUNT

```

299 a      169 c      146 g      259 t      127 others
      /note="Genoscope sequence ID : C0AG167BB11

```

Query Match	7.88;	Score 39.2;	DB 12;	Length 1000;
Best Local Similarity	42.1%;	Pred. No. 4.1;		
Matches 61;	Conservative 24;	Mismatches 60;	Indels 0;	Gaps 0

QY 2 GGGCCGGGGATGGGGCCCGGTGCTTGCACAGGTCGCAAGTGTTCCTCAATTC 61
 || |||| |||| || | : : : : :
 Db 945 GGG 886

Dd

```
97      GAAAGGCCCCCTCTGCCCCCTCCCACATCTGTCGGCGTGGGGGGGGGGGGGGGGGGGG 120
        | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | | | | |
885 CCCCCCCCCCCCCCCCCCCCCCCTTTTATTGTATGSBTSIGGGKGTGGGGGGGG 820
```

02	122	TCACCCCTCCAGGCTTCGCTCTTC	14
		: : : : : :	
Db	825	GKKKCYSSCBCTTTTSTTSSTATC	801

RESULT 7	LOCUS	461 bp	mRNA	linear	EST 16-NOV-2001
BB785045	BB785045				

NE CDNA Mus musculus CDNA clone G430100E18 3', mRNA sequence.	
ACCESSION	BB785045
VERSION	BB785045.1 GI:16953541

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 461)

Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku, Akeiwa, S., Tanaka, T., Tomaru, A., Toyota, T.

TITLE	JOURNAL
RIKEN Encyclopedia of Mouse Full-Length cDNAs (Akimura, T., et al 2001)	Unpublished (2001)

Laboratory for Genome Exploration Research Group, RIKEN Genomic

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,

Normalisation and subtraction of cap-trapper-selected cDNAs to

genes. *Genome Res.* 10 (10), 1617-1630 (2000)

waji, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system-384-format

ratest.eng.uiowa.edu.

TAG_LIB=01-R-B51

TAG_TISSUE=embryo at 13 dpc

TAG_SEQ=AAATCC"

BASE COUNT 60 a 164 c 117 g 188 t

Query Match 7.5%; Score 37.4; DB 9; Length 529;

Best Local Similarity 53.8%; Pred. No. 11; Mismatches 66; Indels 0; Gaps 0;

Matches 77; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Db 64 AGCGCCCTCTGCCCCCGCAATCTGCTCGGGGGTGGGGGGTGGGGGGTGC 123
 Db 160 AGCGCTGGGAGAGCTGCGCCCTGACCTTTCGAGCTGAGAGGTTTCTGGGGGCTTC 219
 Db 124 ACCTCTCAGGTTCTGTTCTTCAAACTTTTGAACCTTAATTGGTGGCTGTAGTGG 183
 Db 220 TCAGACGGTCTTCTTCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 279
 Db 184 GCTCTGAGACTCGGCTCTCTA 206
 Db 280 GCTCTGAGACTCGGCTCTCTA 302

RESULT 10 658 bp mRNA linear EST 23-DEC-1998
 A1322516 m62608.y1 Soares mouse placenta 4NDMP13.5 14.5 Mus musculus cDNA

DEFINITION clone IMAGE:455558 5' similar to gb:x71137 SOX-9 PROTEIN (HUMAN);,
 mRNA sequence.

ACCESSION A1322516 GI:4056945
 VERSION A1322516
 KEYWORDS EST

SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 658)

AUTHORS Marrer,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Scheinberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Wieston,R.

TITLE The WashU-HMNI Mouse EST Project
 JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:772446

This read is a RESEQUENCE of a previously sequenced mouse clone
 IMAGE Consortium (info@image.lnl.gov) for further information.
 correct orientation)
 Seq primer: 40RP from Gibco
 High quality sequence stop: 395.

FEATURES
 source
 Location/Qualifiers
 1..658

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:455558"
 /clone_lib="Soares mouse placenta 4NDMP13.5 14.5"
 /sex="unknown"
 /tissue_type="placenta"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: Placenta; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;

BASE COUNT 122 a 240 c 217 g 77 t 2 others

Query Match 7.5%; Score 37.4; DB 9; Length 558;

Best Local Similarity 87.2%; Pred. No. 12; Mismatches 6; Indels 0; Gaps 0;

Matches 41; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 454 CGAGAGACATCATGAGTGGAGAGCCCGGCTCTGCTGATT 500

Db 1 CGAGAGACATCATGAGTGGAGAGCCCGGCTCTGCTGATT 47

RESULT 11 777 bp DNA linear GSS 03-JUN-1999
 CNS009MF/c

DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC #
 BACR20M21 of RPCT-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION AL054018 GI:4935595
 VERSION AL054018
 KEYWORDS GSS.

SOURCE fruit fly.
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 777)

REFERENCE Direct Submission Genoscope - Centre National de Sequencage :
 Submitted (02-JUN-1999) BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

TITLE Web : www.genoscope.cns.fr
 JOURNAL - Web : www.genoscope.cns.fr

COMMENT The BACP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BACP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osogawa and
 Aaron Mammotser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCT-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDPG from the
 isogenic strain y2; cn bw sp, the same strain used for the BDPG's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 Location/Qualifiers
 1..777

/organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCT-98"
 /clone="BACR20M21"
 /note="end : 77"

BASE COUNT 116 a 128 c 154 g 237 t 142 others

ORIGIN

Query Match 7.5%; Score 37.4; DB 12; Length 777;
 Best Local Similarity 46.7%; Pred. No. 12; Mismatches 37; Indels 0; Gaps 0;

Matches 49; Conservative 19; Mismatches 37; Indels 0; Gaps 0;

Db 42 AAAGTGTCTTTTAATTCGAGAGCCCGCTGCGCCCTCCCAATCTGCTGCTGC 101

Db 746 ARAKTTTGGCAAGCTCGDSGGAGCCCGCCCGCCCGCCCGCCCGGGGGG 687

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
CNS00091P	925 bp DNA linear	SSS 03-JUN-1999								
BACR19D16	of RPEC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.									
AL053013										
AL053013.1	GI:4934461									
SSS.										
fruit fly.										
Drosophila melanogaster										
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.										
1 (bases 1 to 925)										
Genoscope.										
Direct Submission										
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage										
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr										
- web : www.genoscope.cns.fr)										
Determination of this BAC-end sequence was carried out as part of a										
collaboration with the Berkeley Drosophila Genome Project (BDGP).										
The BDGP is constructing a physical map of the Drosophila										
melanogaster genome using these BACs. For further information										
please see http://www.fruitfly.org The BDGP Drosophila										
melanogaster BAC library was prepared by Kazutoyo Osoegawa and										
Aaron Hamster in Pieter de Jong's laboratory in the Department of										
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,										
N.Y. The library is named RPEC1-98 and was constructed by partial										
EcoRI digestion of Drosophila DNA provided by the BDGP from the										
isogenic strain y2; cn bw sp, the same strain used for the BDGP's										
p1 and EST libraries. A more detailed description of the library										
and how to order individual BAC clones, the entire library, or										
filters for hybridization from the BACPAC Resource Center can be										
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .										
location/Qualifiers										
1.925										
/organism="Drosophila melanogaster"										
/db_xref="taxon:7227"										
/clone_lib="RPEC1-98"										
/clone="BACR19D16"										
/note="end : TET3"										
BASE COUNT	120 a	61 c	61 g	172 t	511 others					
ORIGIN										
Query Match	7.4%	Score 37;	DB 12;	Length 925;						
Best Local Similarity	13.1%	Pred. No. 16;								
Matches 44;	Conservative 153;	Mismatches 139;	Indels 1;	Gaps 1;						
QY 160 CCTAATGTGTCCTCTGAGTGGCCCTCTGTGACTCCCGCTCTTAAGTACTTTACC 219										
Db 910 CSSSMSTSSSSBSCSSSSBSSSTSTSMSSBSSSSSSSSSSSGTSSACVCKMASS 851										
QY 220 ACCTACTAGAGCCMAAGAGAGGGCGTGGGAGTAACGMAAGGGCTCCCAACTTTTTTTT 279										
Db 850 SCGGCGCGMABCMCMSSSSSCGSSAAGVAVRASGAGKRGGSAGSASHSSSACBS 791										
QY 280 CCAAGCCAGGCGCAAGCGGGGCTCGGTAAAGTATGGCCAGGGCGCATACGACCACTCT 339										
Db 790 SSSSCASCMASASSSSASSSRSRGGAGAGSASSSSSSSSSASAGVSASASSSSS 731										
QY 340 CAATCACGGGTCTCCGGGTTCCGAGGGCGGACCAAGCCCAACCCCGGGCAATCCGAG 399										
Db 730 SCSSSVSCSSVASMSCSBSSSSASASSSSS-SSSASCASCCCTCTSMSCSCTASMS 672										

LOCUS	BR376636	196 bp	MRNA	linear	EST 24-NOV-2000
DEFINITION	CM3-120068-140800-286-g07 TN0068 Homo sapiens cDNA, mRNA sequence.				
ACCESSION	BR376636				
VERSION	BR376636.1	GI:11338661			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 196)				
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R., Nagai,M.A., da Silva,W Jr., Zaio,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bahia,G.S., Simpson,D.H., Brustin,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,I.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM3&t2=CM3-TN0068-140800-286-g07&t3=2000-08-14&t4=1) Seq primer: puc 18 forward High quality sequence stop: 196. Location/Qualifiers 1. 196 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="TN0068" /dev_stage="Adult" /note="Organ: testis;normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196/716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."				
BASE COUNT	32 a 67 c 28 g 69 t				
ORIGIN					
Query Match	7 4%; Score 36.8; DB 10; Length 196;				
Best Local Similarity	59.6%; Fred. No. 14;				
Matches 62; Conservative	0; Mismatches 42; Indels 0; Gaps 0;				
OY	123 CACCTCCTCAGGTTTCCTTTCAAACTTTTGAACCCCTAATTGGTGCCCTCGAGTG 182 				
Db	93 CACACCTCAGGTTTCCTCCTGCACACCTTGAGACGAGCCTTATGTGGAAGAGCTGTGTG 152				
OY	183 GGCCTGAGTCCGCGCTCCTAGTAACTTTTTCACGACGAC 226 				
Db	153 TTTCTGTGAACCTCCTCCTCCTACACACATACCACTCTC 196				

Search completed: October 10, 2002, 20:28:03

Job time : 702.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 84.9333 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101A-11_COPY_1_500

Perfect score: 500
Sequence: 1 tggccgggggagtgagggcc.....ccgggggtctctctgattt 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq.032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/hold-geneseq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	2509	22	AAD11118	Human small cell 1
2	121.8	24.4	2378	23	AA87738	DNA encoding novel
3	41.8	8.4	506	22	AA181822	Human polynucleoti
4	38.2	7.6	1468	22	AB00613	Human bone marrow
5	36.2	7.2	379	22	AA192147	Human polynucleoti
6	35.4	7.1	114955	20	AA53491	Human adenosine A1
7	34.4	6.9	686	22	AA559273	Human immune/haema
8	34.2	6.8	637	22	AA564084	Human prostate CDN
9	32.6	6.5	1714	24	AB199331	Mouse ischaemic co

10	32.4	6.5	554	22	AA116774	Human breast cancer
11	32.4	6.5	16106	22	AAK83468	Human immune/haema
12	32.4	6.5	16161	22	AAK83469	Human immune/haema
13	32.2	6.4	1556	18	AA194535	Maize cinnamoyl Co
14	32.2	6.4	1177	22	AAH94477	Human foetal cDNA,
15	31.8	6.4	1623	22	AAH30438	Rice apoptosil ind
16	31.6	6.3	994	22	AA115399	Human breast cancer
17	31.2	6.2	646	22	AA564368	DNA encoding novel
18	31.2	6.2	12710	22	AA114807	Human nervous syst
19	31.2	6.2	985	22	AA115317	Human breast cancer
20	31	6.2	21888	22	AAK73863	Human immune/haema
21	30.8	6.2	187	21	AAK11389	Human secreted pro
22	30.8	6.2	387	22	AA559169	Human cancer relat
23	30.8	6.2	1950	20	AAH84531	Human secreted pro
24	30.8	6.2	1950	22	AAH83514	Human secreted pro
25	30.8	6.2	2033	21	AAH08502	Human ISRE-binding
26	30.8	6.2	2476	22	AA531013	Human diagnostic a
27	30.8	6.2	4500	16	AA086237	Megakaryocyte stem
28	30.8	6.2	4548	16	AA086236	Megakaryocyte stem
29	30.8	6.2	4758	22	AAK83830	Human immune/haema
30	30.8	6.2	16329	22	AAH14411	Human nervous syst
31	30.6	6.1	1559	20	AAK25211	Maize cinnamoyl-Co
32	30.6	6.1	2202	22	AAK08650	Human N-copine hom
33	30.6	6.1	3481	23	AB121297	Drosophila melanog
34	30.6	6.1	11470	23	AB121296	Drosophila melanog
35	30.6	6.1	14175	22	AA527814	DNA encoding novel
36	30.6	6.1	14175	22	AAK78859	Human immune/haema
37	30.4	6.1	1278	16	AA080198	Klebsiella sp. nit
38	30.4	6.1	6779	22	AA559570	Protonibacterium
39	30.2	6.0	2568	6	AAH50493	Sequence of skc st
40	30.2	6.0	20445	24	AA519905	Reference sequence
41	30	6.0	811	22	AAH06402	Human cDNA clone (
42	30	6.0	1520	22	AAH68564	Human protein HPO3
43	30	6.0	2078	23	AB111513	Drosophila melanog
44	30	6.0	2798	22	AAH16206	Human cDNA sequenc
45	30	6.0	4216	23	AB111512	Drosophila melanog

ALIGNMENTS

RESULT 1						
AD11118	ID	AD11118	standard; DNA: 2509 BP.			
XX	AD11118:					
XX	24-SEP-2001	(first entry)				
DE	XX	Human small cell lung cancer associated gene, SOX3.				
XX	XX	Human: small cell lung cancer; therapy: hCAP; nucleic acid; NA:				
KW	XX	melanoma; cancer; colon; breast; head; neck; transitional cancer;				
KW	XX	leiomyosarcoma; synovial sarcoma; cytostatic; SOX3; ds.				
OS	XX	Homo sapiens.				
XX	XX					
FT	Key	Location/Qualifiers				
FT	CDS	441..1773				
FT		/*tag= a				
FT		/product= "Human SOX3 protein"				
PN	XX					
XX	XX	WO20015349-A2.				
PD	XX	26-JUL-2001.				
XX	XX					
PF	XX	19-JAN-2001; 2001WO-US02015.				
XX	XX					
PR	XX	21-JAN-2000; 2000US-0489101.				
XX	XX					
PA	XX	(LUDWIG) LUDWIG INST CANCER RES.				
PA	XX	(SLOK) SLOAN KETTERING INST CANCER RES.				
PA	XX	(CORR) CORNELL RES FOUND INC.				

XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX
XX WPI: 2001-457597/49.
DR P-PSDB; AAE05813.

XX Isolated polypeptide, used to treat or prognose a disorder
PT characterized by expression of a hCAP e.g. cancer, is encoded by an
PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -

PS Claim 57; Page 98-99; 152pp; English.

XX The invention relates to nucleic acids and encoded polypeptides which
CC are cancer associated antigens expressed in patients afflicted with
CC small cell lung cancer. The molecules provided by the invention can be
CC used in the diagnosis, monitoring, research or treatment of conditions
CC characterised by the expression of one or more cancer associated
CC antigens. The polypeptide is used to treat a disorder characterised by
CC expression of a hCAP, and determine regression, progression or onset
CC of a condition characterised by expression of an abnormal amount of a
CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
CC are small and non-small cell lung cancer, melanoma, colon, breast, head
CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
CC The present sequence is a small cell lung cancer associated gene
CC designated as NY-SCLC-9, encoding human SOX3 protein.
XX

SQ Sequence 2509 BP; 456 A; 818 C; 757 G; 478 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 2509;
Best Local Similarity 100.0%; Pred. No. 2.9e-143;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGGGGATGGGGGCGGCTGCTGACAGGTGCAAGTGTCTTAATC 60
DB 1 TGGCGGGGATGGGGGCGGCTGCTGACAGGTGCAAGTGTCTTAATC 60
QY 61 CGAAGCGCCCTCTGCCCCCTCCCAATCTGCTGCGGGGATGGGGG 120
DB 61 CGAAGCGCCCTCTGCCCCCTCCCAATCTGCTGCGGGGATGGGGG 120
QY 121 GTCACCTCTCAGAGTTCGTTCTTCAAACTTTTGAACCTTAATGGTGGCTCTGAG 180
DB 121 GTCACCTCTCAGAGTTCGTTCTTCAAACTTTTGAACCTTAATGGTGGCTCTGAG 180
QY 181 TGGGCTCGGAGCTCCCGCTCTTAAGTACCTTACAGCTCTAGGCGAAGAGG 240
DB 181 TGGGCTCGGAGCTCCCGCTCTTAAGTACCTTACAGCTCTAGGCGAAGAGG 240
QY 241 GCGTGGGGTGAAGAAAGGCTCCCACTTTTTCACGCAAGCGGCGG 300
DB 241 GCGTGGGGTGAAGAAAGGCTCCCACTTTTTCACGCAAGCGGCGG 300
QY 301 TCGGTAATGATTTGGCCAGGCGCATCATCGAAGCTGTCAATCAAGGGTCTCCGGGTT 360
DB 301 TCGGTAATGATTTGGCCAGGCGCATCATCGAAGCTGTCAATCAAGGGTCTCCGGGTT 360
QY 361 GCGAGGGGGGAGGAGGCGGCGGCAAGCTCGAGGATTAATTAAGGGGCGGCGAGC 420
DB 361 GCGAGGGGGGAGGAGGCGGCGGCAAGCTCGAGGATTAATTAAGGGGCGGCGAGC 420
QY 421 TAGAGCCAGGCAAGCTGATGAGACTGTTTCGAGAGACTCATCATCAGTCCGAAGC 480
DB 421 TAGAGCCAGGCAAGCTGATGAGACTGTTTCGAGAGACTCATCATCAGTCCGAAGC 480
QY 481 CCGGGGTTCTGCTGATTT 500
DB 481 CCGGGGTTCTGCTGATTT 500

RESULT 2

AA887738 standard; cDNA: 2378 BP.
ID AA887738
XX

AC AA887738;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #23542.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR P-PSDB; ABG23551.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1; SEQ ID No 23542; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 2378 BP; 590 A; 708 C; 664 G; 416 T; 0 other;

Query Match 24.4%; Score 121.8; DB 23; Length 2378;
Best Local Similarity 98.4%; Pred. No. 3.2e-27;
Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 279 TCCAGCAGGCGGAAGGCGGCTCGTAATGATTTGCCAGGCGCATCTGCAAGCTTG 338
DB 444 TGGTGCAGGCGGAAGGCGGCGCTCGTAATGATTTGCCAGGCGCATCTGCAAGCTTG 503
QY 339 TCATACAGGCTCTCGGATTCGAGGCGGAGCAAGCCCAAGCCCGGGGAATCCGA 398
DB 504 TCATACAGGCTCTCGGATTCGAGGCGGAGCAAGCCCAAGCCCGGGGAATCCGA 563
QY 399 GCAAG 403
DB 564 GCAAG 568

AA192147	AA192147 standard; cDNA; 379 BP.
ID	AA192147
AC	AA192147;
XX	
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 12207.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukemia;
KW	nervous system disorders; arthritis; inflammation; ss.
OS	Homo sapiens.
XX	
PN	WO200164835-A2.
XX	
PD	07-SEP-2001.
XX	
PE	26-FEB-2001; 2001WO-US04927.
XX	
PR	28-FEB-2001; 2000US-0515126.
PR	18-MAY-2000; 2000US-0577409.
XX	
PA	(HYSEQ-) HYSEQ INC.
PI	Tang YT, Liu C, Drmanac RT;
DR	WPI; 2001-514838/56.
DR	P-PSDB; AAO12216.
XX	
PT	Isolated nucleic acids and polypeptides, useful for preventing
PT	diagnosing and treating e.g. leukemia, inflammation and immune
PT	disorders -
XX	
PS	Claim 1; SEQ ID NO 12207; 1399pp + Sequence Listing; English.
XX	
CC	The invention relates to human polynucleotides (AA179941-AA193841) and
CC	the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
CC	
SEQ	Sequence 379 BP; 99 A; 97 C; 81 G; 102 T; 0 other;
Query Match	7.2%; Score 36.2; DB 22; Length 379;
Best Local Similarity	56.2%; Pred. No. 0.32;
Matches	66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
OY	55 AATTCGGAAGAGCCCTCTGCCCCCTCCCAATCTGCTGGCGGCGGGGGGT 114
Db	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
245	AAATCTGGGCGGGGACCCGCCGCCCCCTCCGAAACCTCTCTCGGGGCGCTGGG 304
OY	115 GGGGGGGTCACTCCACAGTTTCGTTCTTTAAACTTTTGAACCTTAATGGTGGCC 174
Db	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
305	GGCAGCCACACCCCTTTGGGTGTTTTTCACAACATTTTAAGTCCGGGTTTTCGCC 364
OY	175 T 175
Db	365 T 365

	AAX53491/C	ID AAX53491 standard; DNA; 114955 BP.
XX AC	AAX53491:	
XX DT	05-JUL-1999	(first entry)
DE XX	Human adenosine A1 receptor antisense oligonucleotide fragment.	
KM KM	Antisense oligonucleotide; multiple target; antisense treatment;	
KM KM	Impaired respiration; inflammation; lung disease;	
KM KM	Pulmonary vasoconstriction; inflammation; allergic rhinitis;	
KM KM	Acute asthma; allergy; asthma; impeded respiration;	
KM KM	Respiratory distress syndrome; pain; cystic fibrosis;	
KW KW	Pulmonary hypertension; pulmonary vasoconstriction; emphysema;	
KM KM	Chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;	
KM KM	Colon cancer; breast cancer; lung cancer; pancreatic cancer;	
KN KM	Hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;	
XX OS	Synthetic.	
PN PN	MOS91386-AI.	
PD PD	25-MAR-1999.	
PE PE	17-SEP-1998; 98MO-US19419.	
PR PR	09-JUN-1998; 98US-0093972.	
PR PA	17-SEP-1997; 97US-0059160.	
PA PL	(UYEC-) UNIV EAST CAROLINA.	
PL PT	Nyce JW;	
PT DR	WPJ; 1999-229400/19.	
DR XX	New antisense oligonucleotides used in treatment of, e.g., pulmonary	
XX PT	vasoconstriction	
PS PS	Disclosure; Page 37; 120pp; English.	
CC CC	The specification describes antisense oligonucleotides (AAX52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the junction-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences AAX55272-74. These multiple target oligonucleotides (specifically AAX55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, pneumonia, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, hepatocellular carcinoma, kidney cancer, melanoma, gastric metatases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.	
XX SQ	Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;	
Query Match	7.1% Score 35.4; DB 20; Length 114955;	
Best Local Similarity	32.9%; Pred. No. 3.9;	
Matches 68; Conservative 23; Mismatches 116; Indels 0; Gaps 0;		
CY 280	CGACGCCAGGCGCAAGGGGTCTTGATTCATTGCACGAGGCGCATCTCCGAACACTGT 339 CCACCACNNNNNNNSGGGCCCVAAGNHHNNNSGGGCCVAGANNHNNNSGGGCCVVAG 109514	

OY	340	CATACACGGGTCCTCCGGGGTTGGAGGAGGCGACCAACCCCAACCCGGGGATTCGAG	399
Db	109513	CNNHHNNNSGGGGCCCVAGGCGNNHHNNNSGGGGCCCVAGGCGANNHHNNNSGGGGCCCVAGG	109454
OY	400	CAGGTATATATAGAGGGGGCCCACTGACACCCAGCAGACACTGTGATGGACCTTTCGAGAG	459
Db	109453	CGAGNNHHNNNSGGGGCCCVAGGCGACNNHHNNNSGGGGCCCVAGGCGACGACNNHHNNNSGG	109394
OY	460	AACTCATCATGCTCCGACGAACCCCGCG	486
Db	109393	GGCCCVAGGCGCCAGCCANNHHNNNSGGGG	109367
RESULT 7			
AAK59273			
ID	AAK59273	standard; cdna; 686 bp.	
AAK59273;			
XX	AAK59273;		
XX	06-NOV-2001	(first entry)	
XX			
XX	Human	Immune/haematopoietic antigen encoding cdna SEQ ID NO:4333.	
XX	Human	Immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KM		cytostatic; gene therapy; Vaccine; metastasis; ss.	
XX			
XX	Homo sapiens.		
XX			
PN	WC200157182-A2.		
PD	09-AUG-2001.		
PE			
PE	17-JAN-2001;	2001WO-US01354.	
PR	31-JAN-2000;	2000US-0179065.	
PR	04-FEB-2000;	2000US-0180628.	
PR	24-FEB-2000;	2000US-0184954.	
PR	02-MAR-2000;	2000US-0186350.	
PR	16-MAR-2000;	2000US-0189674.	
PR	17-MAR-2000;	2000US-0190076.	
PR	18-APR-2000;	2000US-0198123.	
PR	19-MAY-2000;	2000US-0205115.	
PR	07-JUN-2000;	2000US-0209467.	
PR	28-JUN-2000;	2000US-0214886.	
PR	30-JUN-2000;	2000US-0215135.	
PR	07-JUL-2000;	2000US-0216647.	
PR	07-JUL-2000;	2000US-0216880.	
PR	11-JUL-2000;	2000US-0217487.	
PR	11-JUL-2000;	2000US-0217496.	
PR	14-JUL-2000;	2000US-0218290.	
PR	26-JUL-2000;	2000US-0220963.	
PR	14-AUG-2000;	2000US-0224518.	
PR	14-AUG-2000;	2000US-0224519.	
PR	14-AUG-2000;	2000US-0225213.	
PR	14-AUG-2000;	2000US-0225214.	
PR	14-AUG-2000;	2000US-0225267.	
PR	14-AUG-2000;	2000US-0225268.	
PR	14-AUG-2000;	2000US-0225270.	
PR	14-AUG-2000;	2000US-0225447.	
PR	14-AUG-2000;	2000US-0225757.	
PR	14-AUG-2000;	2000US-0225758.	
PR	18-AUG-2000;	2000US-0226279.	
PR	22-AUG-2000;	2000US-0226861.	
PR	22-AUG-2000;	2000US-0226868.	
PR	22-AUG-2000;	2000US-0227182.	
PR	30-AUG-2000;	2000US-0227809.	
PR	01-SEP-2000;	2000US-0228287.	
PR	01-SEP-2000;	2000US-0229343.	
PR	01-SEP-2000;	2000US-0229344.	
PR	01-SEP-2000;	2000US-0229345.	
PR	01-SEP-2000;	2000US-0229346.	
PR	01-SEP-2000;	2000US-0229347.	
PR	01-SEP-2000;	2000US-0229348.	
PR	01-SEP-2000;	2000US-0229349.	
PR	01-SEP-2000;	2000US-0229350.	
PR	01-SEP-2000;	2000US-0229351.	
PR	01-SEP-2000;	2000US-0229352.	
PR	01-SEP-2000;	2000US-0229353.	
PR	01-SEP-2000;	2000US-0229354.	
PR	01-SEP-2000;	2000US-0229355.	
PR	01-SEP-2000;	2000US-0229356.	
PR	01-SEP-2000;	2000US-0229357.	
PR	01-SEP-2000;	2000US-0229358.	
PR	01-SEP-2000;	2000US-0229359.	
PR	01-SEP-2000;	2000US-0229360.	
PR	01-SEP-2000;	2000US-0229361.	
PR	01-SEP-2000;	2000US-0229362.	
PR	01-SEP-2000;	2000US-0229363.	
PR	01-SEP-2000;	2000US-0229364.	
PR	01-SEP-2000;	2000US-0229365.	
PR	01-SEP-2000;	2000US-0229366.	
PR	01-SEP-2000;	2000US-0229367.	
PR	01-SEP-2000;	2000US-0229368.	
PR	01-SEP-2000;	2000US-0229369.	
PR	01-SEP-2000;	2000US-0229370.	
PR	01-SEP-2000;	2000US-0229371.	
PR	01-SEP-2000;	2000US-0229372.	
PR	01-SEP-2000;	2000US-0229373.	
PR	01-SEP-2000;	2000US-0229374.	
PR	01-SEP-2000;	2000US-0229375.	
PR	01-SEP-2000;	2000US-0229376.	
PR	01-SEP-2000;	2000US-0229377.	
PR	01-SEP-2000;	2000US-0229378.	
PR	01-SEP-2000;		

KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0106528.
XX 24-FEB-2000; 2000US-0184654.
PR 02-MAR-2000; 2000US-0186350.
PR 15-MAR-2000; 2000US-0189370.
PR 17-MAR-2000; 2000US-0190876.
PR 18-MAR-2000; 2000US-0196173.
PR 19-MAR-2000; 2000US-0205513.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0231968.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
PS Disclosure; SEQ ID NO 38280; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 16106 BP; 3275 A; 4801 C; 4361 G; 3669 T; 0 other;
Query Match 6.5%; Score 32.4; DB 22; Length 16106;
Best Local Similarity 56.6%; Pred. No. 17;
Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
OY 29 TTGACGCGGTTCGAAGTGTTCCTAAATTCGAGAGCGCCCTGCCCCCCCCCA 88
DB 1854 TTTCAAGAGTTCCAAATTTGAGTATTAATTCAGCGCCGCGGCTCATGCTGTA 1795
OY 89 ATCTGCTTGCCTCGCGGGGTGGGGGGGTGCACCTCCTCAGG 134
DB 1794 ATCCCGACGACTTTGGGAGGCGCGGCGGATTCACAGATCAGG 1749
RESULT 12
AAK83469/C
ID AAK83469 standard; DNA; 16161 BP.
XX
AC AAK83469;
XX
XX 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38281.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233066.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR

PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 XX Disclosure; SEQ ID NO 38281; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic

CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 16161 BP; 3307 A; 4825 C; 4360 G; 3669 T; 0 other;
 Query Match 6.5%; Score 32.4; DB 22; Length 16161;
 Best Local Similarity 56.6%; Pred. No. 17;
 Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
 QY 29 TTGACAGGGTTGCAAGTTGTTCTTAATCCGAAGCGCCCTCCGCCCTCCCA 88
 DB 1856 TTTCAGAGTTCCTCAATTTGAGTATGAATATAGCGCAGCGGGTGCATGCCCTGTA 1797
 QY 89 ATCTGCTTCGTCGGGGTGGGGGCTGCGCTCACCTCTCAGG 134
 DB 1796 ATCCAGCAGCATTGGGAGCGGAGTGGCGATCAGATCAGG 1751
 RESULT 13
 AAT94535
 ID AAT94535 standard; cDNA to mRNA; 1556 BP.
 AC AAT94535;
 XX 27-FEB-1998 (first entry)
 DT
 XX
 DE Maize cinnamoyl CoA reductase coding sequence.
 XX
 KW Alfaifa; cinnamoyl CoA reductase; eucaalyplus; transgenic plant; maize;
 RM ligin biosynthesis; ds.
 XX
 OS Zea mays.
 XX
 FH Key Location/Qualifiers
 FT CDS 195..1310
 FT /*tag= a
 FT /product= cinnamoyl CoA reductase
 XX
 PN W09712982-A1.
 XX
 PD 10-Apr-1997.
 XX
 PF 03-OCT-1996; 96WO-FR01544.
 XX
 PR 03-OCT-1995; 95FR-0011623.
 XX
 PA (CNRS) CENT NAT RECH SCI.
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 XX
 PI Beckert M, Boudet AM, Briat JF, Gamas P, Grima-Pettenati J;
 XX
 DR WPI; 1997-226225/20.
 DR P-PSDB; AAM36041.
 XX
 PT Transformation of plants to reduce or increase ligin content -
 PT using DNA encoding alfaifa or maize cinnamoyl CoA reductase or the
 PT corresponding anti-sense mRNA, e.g. to improve digestibility of
 PT fodder crops
 XX
 PS Claim 1; Page 38-40; 86pp; French.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 18.7333 Seconds
(without alignments)
6556.054 Million cell updates/sec

Title: US-09-489-101a-10_COPY_1_500

Perfect score: 500

Sequence: 1 atccctccgcgttcctccca.....aagcgagaagcattcattt 500

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	422	84.4	1500	US-08-704-398-1	Sequence 1, Appl 1
2	422	84.4	1500	PCT-US95-05966-1	Sequence 1, Appl 1
3	362.6	72.5	666	US-08-896-164-7	Sequence 7, Appl 1
4	354.4	70.9	707	US-08-896-164-1	Sequence 1, Appl 1
5	45.2	9.0	7218	US-08-232-463-14	Sequence 14, Appl 1
6	32	6.4	6124	US-08-213-419B-3	Sequence 3, Appl 1
7	31.8	6.4	29793	US-09-302-812-38	Sequence 38, Appl 1
8	31.8	6.4	29793	US-09-511-477-38	Sequence 38, Appl 1
9	31.6	6.3	519	US-08-503-226B-36	Sequence 36, Appl 1
10	31.6	6.3	519	US-08-721-458B-36	Sequence 36, Appl 1
11	31.6	6.3	541	US-08-503-226B-35	Sequence 35, Appl 1
12	31.6	6.3	541	US-08-721-458B-35	Sequence 35, Appl 1
13	31.6	6.3	3588	US-08-197-792-32	Sequence 32, Appl 1
14	31.6	6.3	3588	US-08-459-850-32	Sequence 32, Appl 1
15	31.6	6.3	3588	US-08-459-214-32	Sequence 32, Appl 1
16	31.6	6.3	8355	US-08-406-030A-23	Sequence 23, Appl 1
17	31.2	6.2	897	PCT-US96-05320A-303	Sequence 303, Appl 1
18	30.8	6.2	5852	US-07-867-106-2	Sequence 2, Appl 1
19	30.4	6.1	3580	US-09-081-345-1	Sequence 1, Appl 1
20	30	6.0	4403765	US-09-103-840A-2	Sequence 2, Appl 1
21	30	6.0	4411529	US-09-103-840A-1	Sequence 1, Appl 1
22	29.8	6.0	11464	US-08-991-840A-2	Sequence 2, Appl 1
23	29.6	5.9	3457	US-08-295-882-1	Sequence 1, Appl 1
24	29.4	5.9	2191	US-08-632-806A-6	Sequence 6, Appl 1
25	29.4	5.9	2192	US-08-273-538A-6	Sequence 6, Appl 1
26	29.4	5.9	3144	US-08-687-916-15	Sequence 15, Appl 1
27	29.4	5.9	3144	US-09-138-614-15	Sequence 15, Appl 1

28	29.4	5.9	11970	US-09-345-217-1	Sequence 1, Appl 1
29	29.4	5.9	152331	US-09-128-155-16	Sequence 16, Appl 1
30	29.2	5.8	627	US-09-385-982-502	Sequence 502, Appl 1
31	29.2	5.8	3396	US-08-974-549A-639	Sequence 639, Appl 1
32	29	5.8	1779	US-09-323-427-1	Sequence 1, Appl 1
33	29	5.8	1779	US-09-323-427-2	Sequence 2, Appl 1
34	29	5.8	2031	US-08-262-220-13	Sequence 13, Appl 1
35	29	5.8	2031	US-08-471-733-13	Sequence 13, Appl 1
36	29	5.8	2031	US-08-468-878-13	Sequence 13, Appl 1
37	29	5.8	2031	US-08-750-494-13	Sequence 13, Appl 1
38	29	5.8	2075	US-08-262-220-3	Sequence 3, Appl 1
39	29	5.8	2075	US-08-471-733-3	Sequence 3, Appl 1
40	29	5.8	2075	US-08-468-878-3	Sequence 3, Appl 1
41	29	5.8	2075	US-08-750-494-3	Sequence 3, Appl 1
42	28.8	5.8	1086	US-08-828-009-1	Sequence 1, Appl 1
43	28.8	5.8	1408	US-08-545-196B-11	Sequence 11, Appl 1
44	28.8	5.8	2763	US-09-030-335-3	Sequence 3, Appl 1
45	28.6	5.7	188	US-09-257-541-4	Sequence 4, Appl 1

ALIGNMENTS

RESULT 1
US-08-704-398-1
Sequence 1, Application US/08704398
Patent No. 5679525
GENERAL INFORMATION:
APPLICANT: Peterson, Michael G
APPLICANT: Henkel, Thomas
TITLE OF INVENTION: EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR
TITLE OF INVENTION: BINDING ASSAY
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/704,398
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/246,977
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET INFORMATION: A-59233/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1500 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1500
US-08-704-398-1
Query Match 84.4%; Score 422; DB 1; Length 1500;

Best Local Similarity 100.0%; Pred. No. 4.9e-124;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 ATGGACACACGAGGAGGCTTGGCCGGGAGAGGAGCCGCTGGCATGCTCCATGCGCTGGG 138
DB 1 ATGGACACACGAGGAGGCTTGGCCGGGAGAGGAGCCGCTGGCATGCTCCATGCGCTGGG 60
QY 139 AATTTGGTGAGCGGCTCCACCTAAACGACTTACTAGGAAAGCTATGCGAAATTTATTTA 198
DB 61 AATTTGGTGAGCGGCTCCACCTAAACGACTTACTAGGAAAGCTATGCGAAATTTATTTA 120
QY 199 AAGAGCGAGGAGGATCAACAGTACTTATCTCATGCAAAAGTTGCACAGAGTATAT 258
DB 121 AAGAGCGAGGAGGATCAACAGTACTTATCTCATGCAAAAGTTGCACAGAGTATAT 180
QY 259 GGAATTTGAAAAAGTTTTTTTCCACCTCTTGTGTATCTTATTTGCGAGCGGATGG 318
DB 181 GGAATTTGAAAAAGTTTTTTTCCACCTCTTGTGTATCTTATTTGCGAGCGGATGG 240
QY 319 AAGAAAAAAGAAAGAAATGGAACGCGATGTTGTTCTGAACAGTCTCAACCGTGT 378
DB 241 AAGAAAAAAGAAAGAAATGGAACGCGATGTTGTTCTGAACAGTCTCAACCGTGT 300
QY 379 GCATTTATTTGGATAGAAATAGTGAACAAAGTACAGAGTAACTGGAAGGAAG 438
DB 301 GCATTTATTTGGATAGAAATAGTGAACAAAGTACAGAGTAACTGGAAGGAAG 360
QY 439 AACTATTGCACAGCCAAACATTTGTATATCTGACTCAGACAGCGAAAGCACTTCATT 498
DB 361 AACTATTGCACAGCCAAACATTTGTATATCTGACTCAGACAGCGAAAGCACTTCATT 420
QY 499 TT 500
DB 421 TT 422

RESULT 2
PCT-US95-05966-1
Sequence 1, Application PC/TUS9505966
GENERAL INFORMATION:
APPLICANT: TULARIK, INC.
TITLE OF INVENTION: EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR
TITLE OF INVENTION: BINDING ASSAY
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05966
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,977
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: FP-59233-PC/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 272799
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 1500 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1500
PCT-US95-05966-1

Query Match 84.4%; Score 422; DB 5; Length 1500;
Best Local Similarity 100.0%; Pred. No. 4.9e-124;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 ATGGACACACGAGGAGGCTTGGCCGGGAGAGGAGCCGCTGGCATGCTCCATGCGCTGGG 138
DB 1 ATGGACACACGAGGAGGCTTGGCCGGGAGAGGAGCCGCTGGCATGCTCCATGCGCTGGG 60
QY 139 AATTTGGTGAGCGGCTCCACCTAAACGACTTACTAGGAAAGCTATGCGAAATTTATTTA 198
DB 61 AATTTGGTGAGCGGCTCCACCTAAACGACTTACTAGGAAAGCTATGCGAAATTTATTTA 120
QY 199 AAGAGCGAGGAGGATCAACAGTACTTATCTCATGCAAAAGTTGCACAGAGTATAT 258
DB 121 AAGAGCGAGGAGGATCAACAGTACTTATCTCATGCAAAAGTTGCACAGAGTATAT 180
QY 259 GGAATTTGAAAAAGTTTTTTTCCACCTCTTGTGTATCTTATTTGCGAGCGGATGG 318
DB 181 GGAATTTGAAAAAGTTTTTTTCCACCTCTTGTGTATCTTATTTGCGAGCGGATGG 240
QY 319 AAGAAAAAAGAAAGAAATGGAACGCGATGTTGTTCTGAACAGTCTCAACCGTGT 378
DB 241 AAGAAAAAAGAAAGAAATGGAACGCGATGTTGTTCTGAACAGTCTCAACCGTGT 300
QY 379 GCATTTATTTGGATAGAAATAGTGAACAAAGTACAGAGTAACTGGAAGGAAG 438
DB 301 GCATTTATTTGGATAGAAATAGTGAACAAAGTACAGAGTAACTGGAAGGAAG 360
QY 439 AACTATTGCACAGCCAAACATTTGTATATCTGACTCAGACAGCGAAAGCACTTCATT 498
DB 361 AACTATTGCACAGCCAAACATTTGTATATCTGACTCAGACAGCGAAAGCACTTCATT 420
QY 499 TT 500
DB 421 TT 422

RESULT 3
US-08-896-164-7
Sequence 7, Application US/08896164
Patent No. 6218521
GENERAL INFORMATION:
APPLICANT: OBARA, Yuichi
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
WITH GASTRIC CANCER AND METHODS FOR
TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/896,164
FILING DATE: July 17, 1997
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: No. 6218521man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-896-164-7

Query Match 72.5%; Score 362.6; DB 4; Length 666;
Best Local Similarity 98.9%; Pred. No. 2.1e-105;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 132 GCCTGGGAATTTGGTGAAGCGGCTCCACCTTAACGACTTACTAGGAGAGCTATGCGGAA 191
DB 31 GACAGGGAATTTGGTGAAGCGGCTCCACCTTAACGACTTACTAGGAGAGCTATGCGGAA 90
QY 192 TTATTTAAAGAGCGAGGGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAA 251
DB 91 TTATTTAAAGAGCGAGGGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAA 150
QY 252 GTCATATGGAATTAAGAAAGGTTTTTTTGGCCACCTCTGTGTATATCTTATGCGGAG 311
DB 151 GTCATATGGAATTAAGAAAGGTTTTTTTGGCCACCTCTGTGTATATCTTATGCGGAG 210
QY 312 CGATGGAAGAAAAAAGAAATGAACGAGATGTTGTTCGAACAAAGTCTCA 371
DB 211 TGAATGGAAGAAAAAAGAAATGAACGAGATGTTGTTCGAACAAAGTCTCA 270
QY 372 ACCGTGCTATTTATTTGGATGAGAAATAGTGACCAAGAAATGACGACTTAACCTGGA 431
DB 271 ACCGTGCTATTTATTTGGATGAGAAATAGTGACCAAGAAATGACGACTTAACCTGGA 330
QY 432 AGGAAGAAGTATTTGACAGCCCAACATTTATATATCTGACTCAGACAGGAAAGCA 491
DB 331 AGGAAGAAGTATTTGACAGCCCAACATTTATATATCTGACTCAGACAGGAAAGCA 390
QY 492 CTCATTTT 500
DB 391 CTCATGTT 399

RESULT 4
US-08-896-164-1
Sequence 1, Application US/08896164
Patent No. 6218521
GENERAL INFORMATION:
APPLICANT: OBATA, YUICHI
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR
TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSER: Falek & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/896,164
FILING DATE: July 17, 1997

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6218521man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 707 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-896-164-1

Query Match 70.9%; Score 354.4; DB 4; Length 707;
Best Local Similarity 97.0%; Pred. No. 8.6e-103;
Matches 358; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 132 GCCTGGGAATTTGGTGAAGCGGCTCCACCTTAACGACTTACTAGGAGAGCTATGCGGAA 191
DB 17 GACAGGGAATTTGGTGAAGCGGCTCCACCTTAACGACTTACTAGGAGAGCTATGCGGAA 76
QY 192 TTATTTAAAGAGCGAGGGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAA 251
DB 77 TTATTTAAAGAGCGAGGGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAA 136
QY 252 GTCATATGGAATTAAGAAAGGTTTTTTTGGCCACCTCTGTGTATATCTTATGCGGAG 311
DB 137 GTCATATGGAATTAAGAAAGGTTTTTTTGGCCACCTCTGTGTATATCTTATGCGGAG 196
QY 312 CGATGGAAGAAAAAAGAAATGAACGAGATGTTGTTCGAACAAAGTCTCA 371
DB 197 TGAATGGAAGAAAAAAGAAATGAACGAGATGTTGTTCGAACAAAGTCTCA 256
QY 372 ACCGTGCTATTTATTTGGATGAGAAATAGTGACCAAGAAATGACGACTTAACCTGGA 431
DB 257 ACCGTGCTATTTATTTGGATGAGAAATAGTGACCAAGAAATGACGACTTAACCTGGA 316
QY 432 AGGAAGAAGTATTTGACAGCCCAACATTTATATATCTGACTCAGACAGGAAAGCA 491
DB 317 AGGAAGAAGTATTTGACAGCCCAACATTTATATATCTGACTCAGACAGGAAAGCA 376
QY 492 CTCATTTT 500
DB 377 CTCATGTT 385

RESULT 5
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DOMNER, F.
TITLE OF INVENTION: SCHEIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9PL-F15
; US-08-232-463-14

```

```

Query Match          9.0%; Score 45.2; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 0.00028;
Matches 8; Conservative 202; Mismatches 140; Indels 0; Gaps 0;

```

```

QY 140 AATTGGTGAAGCGCCCTCCACTAAGCACTACTAGGAGCACTGATGCAATATTATTTAA 199
      ||||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1444 AATTGGTGAACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 AAGAGCGAGGGGAGCAACACTATCTATCTGACGAAAGTGCACAGACATATG 259
      RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 125
DB 1384 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 GAATGAAAAAGTTTGGCCCACTCTGCTATATCTATGCGCAGCGATGA 319
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 330 AGAAAAAAGAAACAATGAGCGATGTTGTTGACACAGTCTCAACCGTGTG 379
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1264 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1205
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 380 CATTTATGGATGCAATAGTACCAAGAAATGACAGACGCTAACTGGAAGAA 439
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1204 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1145
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 440 ACTATTCAGACGCCAACAATGTATATATCTGACTCAGACAGCAAG 489
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1144 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
US-08-213-419B-3/C
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inseilburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE REFERENCE: JTI-002CNP
; CURRENT APPLICATION NUMBER: US/08/213,419B
; PRIOR FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
; US-08-213-419B-3

```

```

Query Match          6.4%; Score 32; DB 4; Length 6124;
Best Local Similarity 53.1%; Pred. No. 3.9;
Matches 68; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

```

QY 215 AACAGTACTTATCTTCATGCAAAAGTGCACAGCAAGTCATGGAATGAAAAAGGT 274
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 813 ATATATATATATATATATATAGCAAAATAGATATTTAAATAAAAAGATTACATATAT 754
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 TTTTGGCCCACTCTCTGTATATCTTATGCGCAGCGATGGAAGAAAAAGAAC 334
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 753 AGTTTAAGTAAAGACTGTTAATTCATATTTAAATATTTAAATGAAAAAGAAAAA 694
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 AATGGA 342
      | | | | |
DB 693 AATGAAA 686

```

```

RESULT 7
US-09-302-812-38
; Sequence 38, Application US/09302812B
; Patent No. 633148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG)
; FILE REFERENCE: NIAD 201
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACT
; CURRENT APPLICATION NUMBER: US/09/302,812B
; EARLIER FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; US-09-302-812-38

```

```

Query Match          6.4%; Score 31.8; DB 4; Length 29793;
Best Local Similarity 59.3%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```

```

QY 296 TATATCTTATGGGAGCGGATGAGAAAAAAGAAACAATGGAACCGATGTTGTT 355
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 14612 TATTTGATGAGAGAGATGAGAAAAAAGAGCTTAATAAATGCAATGATATATC 14671
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 CTGACAGAGGCTCAACCGTGCATTTAT 386
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 14672 CTGATTAATAAATCCCTCTTTTTCATCTAAT 14702
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
US-09-511-477-38
; Sequence 38, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.

```


DB 2409 GGGCCAGAAATAGCAAAAAAAAAAGAGACCCCGTGGAGCCCGCTTAAC 2468
QY 367 TCTCAACCGTGTGATTTATTTGGATAGAAATAGTACCAAGAAATGACAGCTAAC 426
DB 2469 CCGGTCCCTTTTAACCAAGTAGGAGGAGGAGTACGACTAAGAACTTTCTTGACA 2528
QY 427 TTGGAGAGAAAGAA 440
DB 2529 GTCCGAGGCCAGAA 2542

RESULT 15
US-08-459-214-32
Sequence 32, Application US/08459214
Patent No. 5716810
GENERAL INFORMATION:
APPLICANT: Anthony J. Mason
INVENTOR: Peter H. Seedling
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or
TITLE OF INVENTION: Beta Chains of Inhibin and Method for Synthesizing Polypeptide
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,214
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/197792
FILING DATE: 17-FEB-1994
PRIOR APPLICATION NUMBER: 07/958414
FILING DATE: 08-OCT-1992
APPLICATION DATA:
APPLICATION NUMBER: 07/74207
FILING DATE: 12-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/215466
FILING DATE: 05-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/906729
FILING DATE: 31-DEC-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/827710
FILING DATE: 07-FEB-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/783910
FILING DATE: 03-OCT-1985
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 297P2D6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 3588 bases
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-459-214-32
Query Match 6.3%; Score 31.6; DB 1; Length 3588;
Best Local Similarity 52.2%; Pred. No. 3.9;
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 307 GGCAGCGATGAGAGAAAAAAAAAGACAATGACAGCGATGTTGTCGACAGAG 366
DB 2409 GGGCCAGAAATAGCAAAAAAAAAAGAGACCCCGTGGAGCCCGCTTAAC 2468
QY 367 TCTCAACCGTGTGATTTATTTGGATAGAAATAGTACCAAGAAATGACAGCTAAC 426
DB 2469 CCGGTCCCTTTTAACCAAGTAGGAGGAGGAGTACGACTAAGAACTTTCTTGACA 2528
QY 427 TTGGAGAGAAAGAA 440
DB 2529 GTCCGAGGCCAGAA 2542

Search completed: October 10, 2002, 18:15:08
Job time: 58.7333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 695.733 Seconds
(without alignments)
9699.805 Million cell updates/sec

Title: US-09-489-101a-10_COPY_1_500

Perfect score: 1 atccccccggttcttccca.....aagcaaacacattttt 500

Sequence:

IDENTITY_NUC

Scoring table: Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estcmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcpl:*
8: em_estcpl:*
9: em_estcpl:*
10: em_estcpl:*
11: em_estcpl:*
12: em_estcpl:*
13: em_estcpl:*
14: em_estcpl:*
15: em_estcpl:*
16: em_estcpl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	478.8	95.4	692	10 BE257118	BE257118 601108855
2	477.2	95.4	692	10 BE257118	BE257118 601108855
3	475.6	95.1	716	10 BE266949	BE266949 601184864
4	475.2	95.0	785	10 BE266949	BE266949 601184864
5	474.6	94.9	690	10 BE266949	BE266949 601184864
6	454.2	90.8	658	10 BE266949	BE266949 601184864
7	423.8	84.8	473	10 BE266949	BE266949 601184864
8	383.4	76.4	473	10 BE266949	BE266949 601184864
9	382.2	76.4	473	10 BE266949	BE266949 601184864
10	382.2	76.4	473	10 BE266949	BE266949 601184864
11	382.2	76.4	473	10 BE266949	BE266949 601184864
12	378.2	75.6	463	10 BE266949	BE266949 601184864
13	373.2	74.6	438	9 AL171575	AL171575 602514159
14	364.2	72.8	799	10 BE266949	BE266949 601184864
15	364.2	72.8	811	10 BE266949	BE266949 601184864
16	364.2	72.8	811	10 BE266949	BE266949 601184864
17	364.2	72.8	944	9 AL532999	AL532999 601184864

18	364.2	72.8	947	9	AL520858	AL520858
19	364.2	72.8	956	9	AL555008	AL555008
20	362.6	72.5	636	9	AM239382	AM239382
21	362.6	72.5	846	9	AL556654	AL556654
22	362.6	72.5	943	10	BE793102	BE793102
23	362.4	72.5	685	10	BC703001	BC703001
24	362.4	72.5	936	9	AL555798	AL555798
25	362.4	72.5	1704	11	BC020780	BC020780
26	361.8	72.4	749	9	AV710541	AV710541
27	360.8	72.2	683	9	AL544062	AL544062
28	360.4	72.1	506	9	AL580717	AL580717
29	360.4	72.0	506	9	AL597064	AL597064
30	355	71.0	634	10	BE242424	BE242424
31	352.2	70.4	574	10	AL599175	AL599175
32	352.2	70.4	617	10	BE791029	BE791029
33	352.2	70.4	778	9	AL546160	AL546160
34	351.4	70.3	839	10	BE623919	BE623919
35	351	70.2	855	9	AL516254	AL516254
36	350	70.0	981	9	AL513760	AL513760
37	350	70.0	984	10	BM472788	BM472788
38	347.8	69.6	947	10	BE783436	BE783436
39	347.4	69.5	810	10	BI837899	BI837899
40	347.2	69.4	569	10	BC828970	BC828970
41	347	69.4	714	10	BC433877	BC433877
42	345.8	69.2	747	10	BI836305	BI836305
43	344.8	69.0	740	10	BE597860	BE597860
44	339	67.8	769	10	BI827542	BI827542
45	338.8	67.8	646	10	BE884239	BE884239

ALIGNMENTS

RESULT 1	BE257118	595 bp	mRNA	linear	EST 13-JUL-2000
LOCUS	601108855F1	NIH_MGC_16	Homo sapiens	CDNA clone	IMAGE:3349813 5'
DEFINITION	BE257118	NIH_MGC_16	Homo sapiens	CDNA clone	IMAGE:3349813 5'
ACCESSION	BE257118	GI:9127588			
VERSION	BE257118	1			
KEYWORDS	EST				
SOURCE	human				
ORGANISM	human				
REPERENCE	human				
AUTHORS	NIH_MGC				
TITLE	NIH_MGC				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: gcrabbs-remail.nih.gov Tissue Procurement: ATCC				

CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: image.lbl.gov
Plate: LINC44 row: m column: 14
High quality sequence stop: 595.
Location/Qualifiers

1. 595
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3349813"
/clone_lib="NIH_MGC_16"
/tissue="type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site: 1; XhoI; Site: 2; EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit

BASE COUNT 185 a 124 c 158 g 128 t
 ORIGIN
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH-MGC library.

Query Match 95.8%; Score 478.8; DB 10; Length 595;
 Best Local Similarity 99.6%; Pred No 1.6e-108;
 Matches 480; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 19 CAGCTCCACGATGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 78
 DB 57 CAGCTCCACGATGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 116
 QY 79 ATGACACACAGGAGGCTCCGCGGAGGAGCGGCTGGCATGCTCATGCGCTGG 138
 DB 117 ATGACACACAGGAGGCTCCGCGGAGGAGCGGCTGGCATGCTCATGCGCTGG 176
 QY 139 AAATTGGTACGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 198
 DB 177 AAATTGGTACGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 236
 QY 199 AAAGAGGAGGAGGATCAACAGTACTATCTCATGCAAAAGTTGACAGAGCATAT 258
 DB 237 AAAGAGGAGGAGGATCAACAGTACTATCTCATGCAAAAGTTGACAGAGCATAT 296
 QY 259 GGAATATGAAAAAGGTTTTTTTGGCCACCTCTGTGTATATCTATGAGGAGCGATG 318
 DB 297 GGAATATGAAAAAGGTTTTTTTGGCCACCTCTGTGTATATCTATGAGGAGCGATG 356
 QY 319 AAGAAAAAAGAAAAAGTGAAGGAGATGTTGTGTGACAGAGTCTCAACCGTGT 378
 DB 357 AAGAAAAAAGAAAAAGTGAAGGAGATGTTGTGTGACAGAGTCTCAACCGTGT 416
 QY 379 GCATTATTTGGGATAGAAATAGTGAACCAAGAAATGACAGAGTCAAACTTGAAGGAAG 438
 DB 417 GCATTATTTGGGATAGAAATAGTGAACCAAGAAATGACAGAGTCAAACTTGAAGGAAG 476
 QY 439 AACTATTGACAGCCAAACATTTGTATATATCTGACTGACAGCAAGGAAACATTCATT 498
 DB 477 AACTATTGACAGCCAAACATTTGTATATATCTGACTGACAGCAAGGAAACATTCATT 536
 QY 499 TT 500
 DB 537 TT 538

RESULT 2
 LOCUS BE253285 662 bp mRNA linear EST 13-JUL-2000
 DEFINITION 60110875F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3349812 5',
 mRNA sequence.
 ACCESSION BE253285
 VERSION BE253285.1 GI:9123350
 KEYWORDS EST.
 SOURCE human
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 662)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: rps@rockefeller.edu
 Tissue: Progenitor
 Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNase Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 High quality sequence stop: 541.
 Location/Qualifiers

source
 1. 662
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3349812"
 /clone_id="NIH_MGC_16"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pORF7; Site: 1; XhoI; Site: 2;
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GACGAGG(6). Library constructed by Ling Hong
 in the Laboratory of Gerald W. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH-MGC library."

BASE COUNT 207 a 143 c 169 g 143 t
 ORIGIN
 Query Match 95.4%; Score 477.2; DB 10; Length 662;
 Best Local Similarity 99.4%; Pred No 4.1e-108;
 Matches 479; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 19 CAGCTCCACGATGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 78
 DB 56 CAGCTCCACGATGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 115
 QY 79 ATGACACACAGGAGGCTCCGCGGAGGAGCGGCTGGCATGCTCATGCGCTGG 138
 DB 116 ATGACACACAGGAGGCTCCGCGGAGGAGCGGCTGGCATGCTCATGCGCTGG 175
 QY 139 AAATTGGTACGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 198
 DB 176 AAATTGGTACGAGGCTCCCAAGGCGCTCAAAAGCGGATACCGGAGCGCTCCG 235
 QY 199 AAAGAGGAGGAGGATCAACAGTACTATCTCATGCAAAAGTTGACAGAGCATAT 258
 DB 236 AAAGAGGAGGAGGATCAACAGTACTATCTCATGCAAAAGTTGACAGAGCATAT 295
 QY 259 GGAATATGAAAAAGGTTTTTTTGGCCACCTCTGTGTATATCTATGAGGAGCGATG 318
 DB 296 GGAATATGAAAAAGGTTTTTTTGGCCACCTCTGTGTATATCTATGAGGAGCGATG 355
 QY 319 AAGAAAAAAGAAAAAGTGAAGGAGATGTTGTGTGACAGAGTCTCAACCGTGT 378
 DB 356 AAGAAAAAAGAAAAAGTGAAGGAGATGTTGTGTGACAGAGTCTCAACCGTGT 415
 QY 379 GCATTATTTGGGATAGAAATAGTGAACCAAGAAATGACAGAGTCAAACTTGAAGGAAG 438
 DB 416 GCATTATTTGGGATAGAAATAGTGAACCAAGAAATGACAGAGTCAAACTTGAAGGAAG 475
 QY 439 AACTATTGACAGCCAAACATTTGTATATATCTGACTGACAGCAAGGAAACATTCATT 498
 DB 476 AACTATTGACAGCCAAACATTTGTATATATCTGACTGACAGCAAGGAAACATTCATT 535
 QY 499 TT 500
 DB 536 TT 537

RESULT 3
 LOCUS BE269949 716 bp mRNA linear EST 13-JUL-2000
 DEFINITION 60118486F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3542655 5',
 mRNA sequence.
 ACCESSION BE269949
 VERSION BE269949.1 GI:9143578
 KEYWORDS EST.
 SOURCE human
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 716)
 NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@b-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMML at: Image.liml.gov
 Plate: L10M234 row: P column: 16
 High quality sequence stop: 042.

FEATURES
 source
 1. 718 sm="Homo sapiens"
 /organism="Homo sapiens"
 /db="GenBank:taxon:9606"
 /clone="IMAGE:3542655"
 /clone.lib="NIH.MGC.8"
 /tissue.type="Burkitt Lymphoma"
 /lab.host="DH10B (phage-resistant)"
 /note="Organ: lymph. Vector: pORF7. Site: 1: XhoI; site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 218 a 155 c 193 g 150 t
 ORIGIN
 Query Match 95.1%; Score 475.6; DB 10; Length 716;
 Best Local Similarity 99.2%; Pred. No. 1e-107;
 Matches 476; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

19 CAGCTCCACGACGCTCCCTCAAGCGGCTCTAAACCGGATACCGAGCGCTCCG 78
 86 CAGCTCCACGACGCTCCCTCAAGCGGCTCTAAACCGGATACCGAGCGCTCCG 145
 79 ATGGACACACGAGGAGGCTCCGCGGAGAGCGGCTCCGATGCTCCGCTGGG 138
 146 ATGGACACACGAGGAGGCTCCGCGGAGAGCGGCTCCGATGCTCCGCTGGG 205
 139 AATTTGTGAGCGGCTCCGACCTAAACGACTTACTAGGAGCTATGCGAAATTATTTA 198
 206 AATTTGTGAGCGGCTCCGACCTAAACGACTTACTAGGAGCTATGCGAAATTATTTA 265
 199 AAGAGCGAGGAGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAGTCATAT 258
 266 AAGAGCGAGGAGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAGTCATAT 325
 259 GGAATATGAAAAAGTTTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 318
 326 GGAATATGAAAAAGTTTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 385
 319 AAGAAAAAAGAAAAAGTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 378
 386 AAGAAAAAAGAAAAAGTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 445
 379 GATTTATGGAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 438
 446 GATTTATGGAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 505
 439 AACTTTGACAGCAGCAAAACATTTGATATATGTAATGTAATGTAATGTAATGTAATG 498
 506 AACTTTGACAGCAGCAAAACATTTGATATATGTAATGTAATGTAATGTAATGTAATG 565
 499 TT 500
 566 TT 567

RESULT 4

B1093196
 LOCUS B1093196 785 bp mRNA linear EST 20-JUN-2001
 DEFINITION M2825137F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4997295 5',
 mRNA sequence.
 ACCESSION B1093196
 VERSION B1093196.1 GI:14511526
 KEYWORDS EST.
 SOURCE Human
 ORGANISM Human
 REFERENCES
 NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 785)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 CONTACT: Robert Strausberg, Ph.D.
 COMMENT
 Email: cga@b-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMML at: http://image.liml.gov
 Plate: L14M1024 row: b column: 16
 High quality sequence stop: 795.
 Location/Qualifiers
 1. 785
 /organism="Homo sapiens"
 /db="GenBank:taxon:9606"
 /clone="IMAGE:4997295"
 /clone.lib="NIH.MGC.10"
 /cell_line="MGC36"
 /lab.host="DH10B"
 /note="Organ: cervix; Vector: pCMV-Sport6; Site: 1: NotI; site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."

BASE COUNT 243 a 168 c 198 g 173 t 3 others
 ORIGIN
 Query Match 95.0%; Score 475.2; DB 10; Length 785;
 Best Local Similarity 99.2%; Pred. No. 1.3e-107;
 Matches 477; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

19 CAGCTCCACGACGCTCCCTCAAGCGGCTCTAAACCGGATACCGAGCGCTCCG 78
 78 CAGCTCCACGACGCTCCCTCAAGCGGCTCTAAACCGGATACCGAGCGCTCCG 137
 79 ATGGACACACGAGGAGGCTCCGCGGAGAGCGGCTCCGATGCTCCGCTGGG 138
 138 ATGGACACACGAGGAGGCTCCGCGGAGAGCGGCTCCGATGCTCCGCTGGG 197
 139 AATTTGTGAGCGGCTCCGACCTAAACGACTTACTAGGAGCTATGCGAAATTATTTA 198
 198 AATTTGTGAGCGGCTCCGACCTAAACGACTTACTAGGAGCTATGCGAAATTATTTA 257
 199 AAGAGCGAGGAGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAGTCATAT 258
 258 AAGAGCGAGGAGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAGTCATAT 317
 259 GGAATATGAAAAAGTTTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 318
 318 GGAATATGAAAAAGTTTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 377
 319 AAGAAAAAAGAAAAAGTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 378
 378 AAGAAAAAAGAAAAAGTTTGGCCACCTCTTGATATCTTATGCGACGCGATGG 437
 379 GATTTATGGAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 438
 438 GATTTATGGAATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 497

```

QY 439 AACATTGACAGCCAAACATGTATATCTGACTCAGACAGCAAGCACTTCATT 498
DB 498 AACATTGACAGCCAAACATGTATATCTGACTCAGACAGCAAGCACTTCATT 557
QY 499 T 499
DB 558 T 558

RESULT 5
BE513818
LOCUS BE513818
DEFINITION 601315659p1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3534191 5',
mRNA sequence.
ACCESSION BE513818
VERSION BE513818.1 GI:9721030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 690)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: L10M238 row: 3 column: 16
High quality sequence stop: 656.
Location/Qualifiers
1..690
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3534191"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 211 a 145 c 184 g 145 t
ORIGIN
Query Match 94.9%; Score 474.6; DB 10; Length 690;
Best Local Similarity 99.2%; Pred. No. 1.8e-107;
Matches 477; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 19 CAGCTCCAGCTACGCTCCCTCAAGCGGCTCTTAACCGGATTAACCGGAGCGCTCCC 78
DB 87 CAGCTCCAGCTACGCTCCCTCAAGCGGCTCTTAACCGGATTAACCGGAGCGCTCCC 146
QY 79 ATGACACACAGGAGGCTGCGCGGAGGAGCGGCTCCATGCTCCATGCTCGCTGG 138
DB 147 ATGACACACAGGAGGCTGCGCGGAGGAGCGGCTCCATGCTCCATGCTCGCTGG 206
QY 139 AAATTGTGAGCGGCTCCCTCAAGCGGCTCTTAACCGGATTAACCGGATTAACCGG 198
DB 207 AAATTGTGAGCGGCTCCCTCAAGCGGCTCTTAACCGGATTAACCGGATTAACCGG 266
QY 139 AAAGAGCGGAGGATCAAAACAGTACTTCTTCTTCAAGAGTTGCAAGAGTCTATG 258
DB 267 AAAGAGCGGAGGATCAAAACAGTACTTCTTCTTCAAGAGTTGCAAGAGTCTATG 326

```

```

QY 259 GGAATGAAAAAGTTTWTGCCCCACCTCTTGTATATCTTATGAGGACGCGATG 318
DB 337 GGAATGAAAAAGTTTWTGCCCCACCTCTTGTATATCTTATGAGGACGCGATG 366
QY 319 AAGAAAAAAGAAAGAAATGAGACGCGATGTTGTTCTGACAAGAGTCTCAACCGTGT 378
DB 387 AAGAAAAAAGAAAGAAATGAGACGCGATGTTGTTCTGACAAGAGTCTCAACCGTGT 446
QY 379 GCATTATGAGATGAGAAATGATGACCAAGAAATGACGAGCTAACTTGAAGGAAG 438
DB 447 GCATTATGAGATGAGAAATGATGACCAAGAAATGACGAGCTAACTTGAAGGAAG 506
QY 439 AACATTGACAGCCAAACATGTATATCTGACTCAGACAGCAAGCACTTCATT 498
DB 507 AACATTGACAGCCAAACATGTATATCTGACTCAGACAGCAAGCACTTCATT 566
QY 499 T 499
DB 567 T 567

RESULT 6
BE269143
LOCUS BE269143
DEFINITION 601184493p1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3542135 5',
mRNA sequence.
ACCESSION BE269143
VERSION BE269143.1 GI:9142758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 658)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: L10M233 row: 3 column: 24
High quality sequence stop: 583.
Location/Qualifiers
1..658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3542135"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 195 a 136 c 182 g 145 t
ORIGIN
Query Match 90.8%; Score 454.2; DB 10; Length 658;
Best Local Similarity 99.0%; Pred. No. 2.1e-102;
Matches 478; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 19 CAGCTCCAGCTACGCTCCCTCAAGCGGCTCTTAACCGGATTAACCGGAGCGCTCCC 78

```

Db 87 CAGTCTCCAGTACGTCCTCCGAAAGCGGCTCTTAAACCCGGATACCGGAGCGCTCCCC 146
 QY 79 ATGACCCACACGAGAGGCTTCCCGCGAGAGAGCCCTCCGCAATGCTCCATGCGCTGGG 138
 Db 147 ATGACCCACACGAGAGGCTTCCCGCGAGAGAGCCCTCCGCAATGCTCCATGCGCTGGG 206
 QY 139 AAATTGGTGAGCGGCTCCACCTAAAGCACTACTAGAGAGAGTATGCAATATATTTA 198
 Db 207 AAATTGGTGAGCGGCTCCACCTAAAGCACTACTAGAGAGAGTATGCAATATATTTA 266
 QY 199 AAGAGCGAGGAGATCAACAGTACTATCTTATCATGCAAAAAGTTGACAGAGATCATAT 258
 Db 267 AAGAGCGAGGAGATCAACAGTACTATCTTATCATGCAAAAAGTTGACAGAGATCATAT 326
 QY 259 GGAATGTAAAAAGGTTTTTTTGGCCACCTCTTGTATATCTTATAGGAGAGGATGG 318
 Db 327 GGAATGTAAAAAGGTTTTTTTGGCCACCTCTTGTATATCTTATAGGAGAGGATGG 386
 QY 319 AAGAAAAAAGAAACAATAGAAAGGATGCTTGTGTAAGAAAGAGTCAACCGTGT 378
 Db 387 AAGAAAAAAGAAACAATAGAAAGGATGCTTGTGTAAGAAAGAGTCAACCGTGT 446
 QY 379 GCATTATTT-GGGATAGGAAATAGTACCAAGAAATGCAAGCACTTAAAGTGAAGAAA 437
 Db 447 GCATTATTTGGGATAGGAAATAGTACCAAGAAATGCAAGCACTTAAAGTGAAGAAA 506
 QY 438 GAATCTATGCAAGCAACATGTTATATCTGACTGACGCAAGCAAGCACTTCAAT 497
 Db 507 GAATCTATGCG-CAGCCAAACATGTTGTATATCTGACTGACGCAAGCAAGCACTTCAAT 565
 QY 498 TTT 500
 Db 566 GTT 568

RESULT 7
 BE269331 524 bp mRNA linear EST 13-JUL-2000
 LOCUS 601185204P1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3542833 5',
 DEFINITION mRNA sequence.
 ACCESSION BE269331
 VERSION BE269331.1 GI:9142950
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 524)
 NIH-MGC http://mgi.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at: image.lmnl.gov
 Plate: LINC235 row: h column: 02
 High quality sequence stop: 524.
 Location/Qualifiers
 1..524
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3542833"
 /clone_lib="NIH_MGC_8"
 /tissue_type="Burkitt Lymphoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lymph. Vector: pOT87; Site: 1. XhoI; Site: 2.
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(g). Size-selected >500bp for average

Insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

Query Match 84.8%; Score 423.8; DB 10; Length 524;
 Best Local Similarity 99.3%; Pred. No. 7.3e-95;
 Matches 436; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

BASE COUNT 158 a 113 c 147 g 106 t
 ORIGIN
 19 CAGTCTCCAGTACGTCCTCCGAAAGCGGCTCTTAAACCCGGATACCGGAGCGCTCCCC 78
 Db 86 CAGTCTCCAGTACGTCCTCCGAAAGCGGCTCTTAAACCCGGATACCGGAGCGCTCCCC 145
 QY 79 ATGACCCACACGAGAGGCTTCCCGCGAGAGAGCCGCTGCAATGCTCCATGCGCTGGG 138
 Db 146 ATGACCCACACGAGAGGCTTCCCGCGAGAGAGCCGCTGCAATGCTCCATGCGCTGGG 205
 QY 139 AAATTGGTGAGCGGCTCCACCTAAAGCACTACTAGAGAGAGTATGCAATATATTTA 198
 Db 206 AAATTGGTGAGCGGCTCCACCTAAAGCACTACTAGAGAGAGTATGCAATATATTTA 265
 QY 199 AAGAGCGAGGAGATCAACAGTACTATCTTATCATGCAAAAAGTTGACAGAGATCATAT 258
 Db 266 AAGAGCGAGGAGATCAACAGTACTATCTTATCATGCAAAAAGTTGACAGAGATCATAT 325
 QY 259 GGAATGTAAAAAGGTTTTTTTGGCCACCTCTTGTATATCTTATAGGAGAGGATGG 318
 Db 326 GGAATGTAAAAAGGTTTTTTTGGCCACCTCTTGTATATCTTATAGGAGAGGATGG 385
 QY 319 AAGAAAAAAGAAACAATAGAAAGGATGCTTGTGTAAGAAAGAGTCAACCGTGT 378
 Db 386 AAGAAAAAAGAAACAATAGAAAGGATGCTTGTGTAAGAAAGAGTCAACCGTGT 445
 QY 379 GCATTATTT-GGGATAGGAAATAGTACCAAGAAATGCAAGCACTTAAAGTGAAGAAA 437
 Db 446 GCATTATTTGGGATAGGAAATAGTACCAAGAAATGCAAGCACTTAAAGTGAAGAAA 505
 QY 438 GAATCTATGCAAGCAACATGTTATATCTGACTGACGCAAGCAAGCACTTCAAT 497
 Db 506 GAATCTATGCGACAGCCAAA 524

RESULT 8
 BE609533 473 bp mRNA linear EST 17-APR-2001
 LOCUS 323418 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION BE609533
 ACCESSION BE609533.1 GI:13659512
 VERSION BE609533.1
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 REFERENCE 1 (bases 1 to 473)
 Fahnenkrug, S.C., Fekling, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
 and Keele, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 Contact: Smith TPJ
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mael.marc.usda.gov

Single pass sequencing. Bases called and all trimmed with phred
 v1.960904.e. Vector identified by cross-match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT


```

DB 366 CTCAGACAGCAAGCACTTCATGTT 392
|||||
RESULT 12
LOCUS BF226018 463 bp mRNA linear EST 29-DEC-2000
DEFINITION uc47c01.y1 NCL-CGAP_Mam6 Mus musculus cDNA clone IMAGE:3672192 5'
similar to SW:RBR_MOUSE_P3126 J KAPPA-RECOMBINATION SIGNAL
BINDING PROTEIN ; mRNA sequence.
ACCESSION BF226018
VERSION BF226018.1 GI:11133663
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 463)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml
MGI:1432960
FEATURES
source High quality sequence stop: 341.
location/Qualifiers
1..463
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="3672192"
/clone_lib="NCL-CGAP_Mam6"
/sex="Female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DHI0B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Saliv;
Site:2: Notti; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 141 a 94 c 124 g 104 t
ORIGIN
Query Match 75.6%; Score 378.2; DB 10; Length 463;
Best Local Similarity 90.4%; Pred. No. 1.5e-83;
Matches 404; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
54 AACCGGATACCGGAGCGCTCCCATGACACACGAGGCGCTTGGCCGAGGAGGC 113
|||||
DB 1 AACCGGATACCGGAGCGCTCCCATGACACACGAGGCGCTTGGCCGAGGAGGC 60
114 GCTGCGCATGCTTCATGCGCTGGGAAATTGGTGAAGCGCTTCAACGACTTTC 173
|||||
DB 61 GCTGCGCATGCTTCATGCGCTGGGAAATTGGTGAAGCGCTTCAACGACTTTC 120
174 TAGGGAAGCTATGCGAATTTATTTAAAGAGGAGGATCAACAGTACTTATCTTCA 233
|||||
DB 121 TAGGGAAGCTATGCGAATTTATTTAAAGAGGAGGATCAACAGTACTTATCTTCA 180
234 TGCAAAAGTTGCACAGAGTATGAAATGAAAGTTTGGCCACCTCTTG 293
|||||
DB 181 TGCAAAAGTTGCACAGAGTATGAAATGAAAGTTTGGCCACCTCTTG 240
294 TGTAATCTTATGGGCGAGCGATGAGAAAGAAAGAAAGAAAGTGAAGCGATGTTG 353

```

```

DB 241 TGTATCTTATGGCGATGTTGAGAGAAAAAGAGCAATGAGACGATGATG 300
|||||
QY 354 TTCGACACAGCTCCCAACCGTCGATTTATGGGATGAGAAATGACCAACAAT 413
|||||
DB 301 TTCTGACAGAGTCTCCACCTGCTTGTTCGATGAGAAACGATGACCAACAAT 360
414 GCAGCAGTAACTTGGAGAGAAAGTATTCACAGCCCAACATGATATATCTCA 473
|||||
DB 361 GCAGCAGTAACTTGGAGAGAAAGTATTCACAGCCCAACATGATATATTCG 420
QY 474 CTCAGACAGCAAGCACTTCATGTT 500
|||||
DB 421 ATCAGACAGCAAGCACTTCATGTT 447
|||||
RESULT 13
LOCUS AA171575 438 bp mRNA linear EST 23-DEC-1996
DEFINITION zp23a04.r1 Stratagene neuroepithelium (#937231) Homo sapiens CDNA
clone IMAGE:610254 5' similar to gb:L07872_cds1 J
KAPPA-RECOMBINATION SIGNAL BINDING PROTEIN (HUMAN); mRNA sequence.
ACCESSION AA171575
VERSION AA171575.1 GI:1750644
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldi, M.F., Chapelli, B.,
Chisoe, S., Dietrich, N., Dubouche, T., Favello, A., Gish, W., Hawkins
M., Hultman, M., Kueba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, J., Prange, C., Rikkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisan, E.,
Underwood, K., Woldman, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280 000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
JOURNAL Contact: Wilson RK
MEDLINE Washington University School of Medicine
COMMENT 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28W13 rev2 from Amersham
High quality sequence stop: 322.
FEATURES
source Location/Qualifiers
1..438
/organism="Homo sapiens"
/db_xref="GDB:462513"
/db_xref="taxon:9606"
/clone_image="610254"
/clone_lib="Stratagene neuroepithelium (#937231)"
/dev_stage="Ntera-2/RA neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site:1: EcoRI; Site:2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2
cells (Ntera 2/cl.D1) induced with Retinoic acid for 24
hours. Average insert size: 1.5 kb; DmL-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGGCGAG 3' -3' adaptor
sequence: 5' CTCGAGTCTTCTTTTCTTTT 3'"
BASE COUNT 137 a 83 c 107 g 102 t 9 others
ORIGIN
Query Match 74.6%; Score 373; DB 9; Length 438;
Best Local Similarity 96.6%; Pred. No. 3e-82;
Matches 399; Conservative 0; Mismatches 11; Indels 3; Gaps 2;
81 GGACACAGGAGGCTTGGCCGAGGAGCGGCTCGCATGCTCGATCGCTGGAA 140
|||||

```

Db 1 GACACACAGGAGGCGCTCCGCGGAGAGAGCCNCTGCGATCTCCATCGCTGGGAA 60
 QY 141 ATTGGTGAGCGGCGCTCCACCTAAAGACTTACTAGGAGAGCTATGCG-AAATATTTAA 199
 Db 61 ATTGGTGAGCGGCGCTCCACCTAAAGACTTACTAGGAGAGCTATGCGNNATTTATTA 120
 QY 200 AAGAGCGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGAGTATATG 259
 Db 121 AAGAGCGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGAGTATATG 180
 QY 260 GAAATGAAAAAGTCTTTTTCGCCACCTCTTGTATATCTTATGCGAGCGAGTGA 319
 Db 181 GAAATGAAAAAGTCTTTTTCGCCACCTCTTGTATATCTTATGCGAGCGAGTGA 240
 QY 320 AGAAAAAGGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGAGTATG 379
 Db 241 AGAAAAAGGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGAGTATG 300
 QY 380 CATTTATGAGTATGAGTAATGACCAAGAAATGACAGAGCT-AAACTTGAAGAGAA 437
 Db 301 CATTTATGAGTATGAGTAATGACCAAGAAATGACAGAGCTTAACTTGAAGAGAA 360
 QY 438 GAACCTTTCAGCAGCAAAACATGATATCTATCTGACAGCAGCAGCAAGC 490
 Db 361 GAACCTTTCAGCAGCAAAACATGATATCTGACAGCAGCAGCAAGC 413

RESULT 14
 AL602969 599 bp mRNA linear EST 14-AUG-2001
 LOCUS DKF60686H1720.1 686 (synonym: h1cc3) Homo sapiens cDNA clone
 DEFINITION DKF60686H1720.5, mRNA sequence.
 ACCESSION AL602969
 VERSION AL602969.1 GI:15166475
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 599)
 AUTHORS Mambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.
 TITLE EST (Mambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Mambutt R

FEATURES
 source
 1..599
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKF60686H1720"
 /clone_1lb="686 (synonym: h1cc3)"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pTrioPlEx2; Site_1: SfiI; Site_2: SfiI;
 cDNA-collection"
 BASE COUNT 198 a 107 c 142 g 152 t

Query Match 72.8%; Score 364.2; DB 9; Length 599;
 Best Local Similarity 99.2%; Pred. No. 4.5e-80;
 Matches 366; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 132 GCCTGGAAATTTGGTAGCGGCTCCACCTAAAGACTTACTAGGAGAGCTATGCGAA 191
 Db 38 GACAGGAAATTTGGTAGCGGCTCCACCTAAAGACTTACTAGGAGAGCTATGCGAA 97
 QY 192 TTTATTTAAAGAGCGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGA 251
 Db 98 TTTATTTAAAGAGCGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGA 157
 QY 252 GTCATATGAAATGAAAAAGTTTTTTCGCCACCTCTTGTATATCTTATGCGAG 311
 Db 158 GTCATATGAAATGAAAAAGTTTTTTCGCCACCTCTTGTATATCTTATGCGAG 217
 QY 312 CGCATGCAAAAAAGGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGA 371
 Db 218 CGCATGCAAAAAAGGAGGAGTCAACAGTACTTATCTTCATGCAAAAGTGCACAGA 277
 QY 372 ACCGTGTGATTTATGCGATGAGAAATGTCACAAAGATGACAGCTTAACTTGA 431
 Db 278 ACCGTGTGATTTATGCGATGAGAAATGTCACAAAGATGACAGCTTAACTTGA 337
 QY 432 AGGAAGACTATTCACACGCCAAACATGATATCTATCTGACAGCAGCAAGAGCA 491
 Db 338 AGGAAGACTATTCACACGCCAAACATGATATCTATCTGACAGCAGCAAGAGCA 397
 QY 492 CTTCATTTT 500
 Db 398 CTTCATGTT 406

RESULT 15
 BG614689 799 bp mRNA linear EST 18-APR-2001
 LOCUS 602642184F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4773165 5',
 DEFINITION mRNA sequence.
 ACCESSION BG614689
 VERSION BG614689.1 GI:1366060
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 799)
 AUTHORS NIH-MGC
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM1644 row: 9 column: 22
 High quality sequence stop: 679.

FEATURES
 source

1..799

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4773165"

/clone_1lb="NIH_MGC_61"

/tissue_type="embryonal carcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgagc); Site_2: SfiI (ggcattatggc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATATGACC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGGAGCGGAGACATG-3' (30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library.
BASE COUNT 252 a 146 c 194 g 207 t
ORIGIN

Query Match 72.8%; Score 364.2; DB 10; Length 799;
Best Local Similarity 99.2%; Pred. No. 4.4e-80;
Matches 366; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 132 GCGTGGGAAATTTGGTGGAGGGGCTCCACCTTAAGCACTTACTAGGGAGCTATGCGAA 191
Db 38 GACAGGGAATTTGGTGGAGGGGCTCCACCTTAAGCACTTACTAGGGAGCTATGCGAA 97
OY 192 TTATTTAAAGAGCGAGGGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAA 251
Db 98 TTATTTAAAGAGCGAGGGGATCAACAGTACTTATCTTCATGCAAAAGTTGCACAGAA 157
OY 252 GTCATATGAAATGAAAGGTTTTTTGGCCACCTCCTGTGTATATCTTATGGGCG 311
Db 158 GTCATATGAAATGAAAGGTTTTTTGGCCACCTCCTGTGTATATCTTATGGGCG 217
OY 312 CCGATGGAAGAAAAAACAATGGAACCGGATGTTGTCTGAACAAGAGTCTCA 371
Db 218 CCGATGGAAGAAAAAACAATGGAACCGGATGTTGTCTGAACAAGAGTCTCA 277
OY 372 ACCGTGTCAATTTATTTGGATAGAAATAGTGAACAAGAAATGCACAGCTAACTTGA 431
Db 278 ACCGTGTCAATTTATTTGGATAGAAATAGTGAACAAGAAATGCACAGCTAACTTGA 337
OY 432 AGGAAAGACTATTTGCACAGCAAAACATTTATATCTGACTGACAGCAAGCAAGCA 491
Db 338 AGGAAAGACTATTTGCACAGCAAAACATTTATATCTGACTGACAGCAAGCAAGCA 397
OY 492 CTTGATTTT 500
Db 398 CTTGATGTT 406
```

Search completed: October 10, 2002, 20:27:56
Job time : 702.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 84.933 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-10_COPY_1_500

Perfect score: 500
Sequence: 1 atccccccggtttctctca.....aagcgaaagcattcattt 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseg_032802.*
2: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseg/geneseg-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	1580	20	AAK39671
2	500	100.0	1580	22	AAK39671
3	432.8	86.6	1581	13	AAK20395
4	422	84.4	1500	17	AAK06947
5	364.2	72.8	1568	21	AAK18333
6	362.6	72.5	666	20	AAK40082
7	354.4	70.9	707	20	AAK40076
8	348.8	69.8	1336	22	AAK22790
9	327.2	65.4	373	22	AAK42783

10	320.8	64.2	1599	20	AAK39661	Breast cancer asso
11	319.6	63.9	473	22	AAK33426	Human colon cancer
12	312.6	62.5	665	22	AAK22796	Human prostate can
13	282.8	56.6	768	22	AAK22792	Human prostate can
14	281.8	56.4	625	22	AAK22795	Human prostate can
15	238.8	47.8	815	22	AAK22791	Human breast can
16	213.8	42.8	391	22	AAK55536	Human breast tumor
17	181.2	36.2	229	21	AAK00786	Human secreted pro
18	171.4	34.3	538	23	AAK92920	Human lung tumour
19	171.4	34.3	538	23	AAK23366	Human lung tumour
20	160.8	32.2	741	22	AAK22793	Human prostate can
21	127.4	25.5	2932	23	AAK06269	Drosophila melanog
22	115.2	23.0	251	19	AAK12556	Human ballelic po
23	115.2	23.0	251	19	AAK12557	Human ballelic po
24	115.2	23.0	262	20	AAK87430	Human single nucle
25	115.2	23.0	262	20	AAK87431	Human single nucle
26	111	22.2	5829	23	ABK06268	Drosophila melanog
27	67	13.4	303	21	AAK08033	Human secreted pro
28	65.2	13.0	1127	22	ABK3044	Human transcriptio
29	65.2	13.0	2370	21	AAK7028	Human ORFX ORF2583
30	47.4	9.5	819	22	AAK22649	Human gastric can
31	47.4	9.5	819	22	AAK22649	Human gastric can
32	41.2	8.2	276	23	AAK1624	DNA encoding novel
33	41.2	8.2	3662	23	AAK1622	DNA encoding novel
34	41.2	8.2	3662	23	AAK2571	DNA encoding novel
35	35.6	7.1	8119	21	AAK35392	Maize starch bran
36	35	7.0	32367	19	AAK35620	Human SHOX (short
37	34.6	6.9	8964	24	AAK35354	Human immune syste
38	34.6	6.9	8964	24	AAK35354	Human immune syste
39	34.4	6.9	1154	20	AAK16396	Human gene regulat
40	34.4	6.9	3148	22	AAK54436	S. epidermidis gen
41	34.4	6.9	3450	22	AAK54587	S. epidermidis gen
42	34	6.8	3769	22	AAK54742	Human immune syste
43	34	6.8	18598	24	ABK2387	Human HBM gene reg
44	33.8	6.8	26928	24	ABK2620	Human gene regulat
45	33.8	6.8	10647	24	AAK1396	Human gene regulat
				21	AAK21288	Human low adenosin

ALIGNMENTS

RESULT 1	
AAK39671	
ID	AAK39671 standard; DNA; 1580 BP.
XX	
AC	AAK39671:
XX	
02-JUL-1999	(first entry)
XX	
XX	Renal cancer associated gene.
DE	
XX	
XX	Cancer associated antigen; diagnosis: research; treatment: human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998: 98WO-US14679.
XX	
XX	22-JUN-1998: 98US-0102322.
PR	17-JUL-1997: 97US-0896164.
PR	10-OCT-1997: 97US-0061599.
PR	10-OCT-1997: 97US-0061765.
PR	10-OCT-1997: 97US-0948705.
PR	11-OCT-1997: 97GB-0021697.
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;


```

|||||
Db 241 GTTGCACAGAGCATATGAAATGAAAAAGTTTTTTGGCCACCTCTGTGTATAT 300
Qy 301 CTTATGGGCGGCGATGAGAGAAAAAACAATGAAACGCGATGCTGTTCGAA 360
Db 301 CTTATGGGCGGCGATGAGAGAAAAAACAATGAAACGCGATGCTGTTCGAA 360
Qy 361 CAGAGCTCTCAACCGTGTGCACTTTATTTGGATGAGAAATGTCACCAAGAAATGACAG 420
Db 361 CAGAGCTCTCAACCGTGTGCACTTTATTTGGATGAGAAATGTCACCAAGAAATGACAG 420
Qy 421 CTTAACTTGAGAGAAAGAACTATTGACAGCCAAAAATGTATATATCTGACTCAGAC 480
Db 421 CTTAACTTGAGAGAAAGAACTATTGACAGCCAAAAATGTATATATCTGACTCAGAC 480
Qy 481 AAGCGAAGCACTTCATTTT 500
Db 481 AAGCGAAGCACTTCATTTT 500

```

RESULT 3
AAQ20595
ID AAQ20595 standard; DNA; 1581 BP.

AC AAQ20595;

DT 15-APR-1992 (first entry)

DE JK recombinant signal sequence binding protein gene.

XX JKRS; immunoglobulin; ss.

XX Key Location/Qualifiers

FT CDS 1..1581

FT CDS /tag= a

PN JP0327283-A.

PD 09-DEC-1991.

XX 26-MAR-1990; 90JP-0077317.

XX 26-MAR-1990; 90JP-0077317.

XX (TOYM) TOYOBO KK.

XX WPI; 1992-029690/04.

XX P-PSDB; AAR20575.

XX JK recombinant signal sequence binding protein - binds

XX specifically on immunoglobulin gene, obtd. by culturing

XX transformant of plasmid, then collecting protein

XX Claim 2; Fig 6; 11pp; Japanese.

XX The sequence is that which encodes JK recombinant signal sequence

XX (JKRS) binding protein. Using the sequence mass production of the

XX protein is possible by genetic engineering. See also AAQ20596-020601.

XX Sequence 1581 BP; 460 A; 380 C; 392 G; 349 T; 0 other;

Qy Query Match 86.6%; Score 432.8; DB 13; Length 1581;
Best Local Similarity 91.6%; Pred. No. 2,7e-125;
Matches 458; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```

Qy 121 CATGCTCCATCGCCCTGGGAAATTTGGTGAAGCGGCTCCACCTAAACGACTTACTAGGAA 180
Db 121 CATGCTCCATCGCCCTGGGAAATTTGGTGAAGCGGCTCCACCTAAACGACTTACTAGGAA 180
Qy 181 GCTATGCGCAATTTATTTAAAGAGCGAGGATCAACAGACTTATCTCATGCAAAA 240
Db 181 GCTATGCGCAATTTATTTAAAGAGCGAGGATCAACAGACTTATCTCATGCAAAA 240
Qy 241 GTTGCACAGAGCTCATATGAAATGAAAAAAGTTTTTGGCCACCTCTGTGTATAT 300
Db 241 GTTGCACAGAGCTCATATGAAATGAAAAAAGTTTTTGGCCACCTCTGTGTATAT 300
Qy 301 CTTATGGGCGGCGATGAGAGAAAAAACAATGAAACGCGATGCTGTTCGAA 360
Db 301 CTTATGGGCGGCGATGAGAGAAAAAACAATGAAACGCGATGCTGTTCGAA 360
Qy 361 CAGAGCTCTCAACCGTGTGCACTTTATTTGGATGAGAAATGTCACCAAGAAATGACAG 420
Db 361 CAGAGCTCTCAACCGTGTGCACTTTATTTGGATGAGAAATGTCACCAAGAAATGACAG 420
Qy 421 CTTAACTTGAGAGAAAGAACTATTGACAGCCAAAAATGTATATATCTGACTCAGAC 480
Db 421 CTTAACTTGAGAGAAAGAACTATTGACAGCCAAAAATGTATATATCTGACTCAGAC 480
Qy 481 AAGCGAAGCACTTCATTTT 500
Db 481 AAGCGAAGCACTTCATTTT 500

```

RESULT 4
AAT06947
ID AAT06947 standard; cDNA; 1500 BP.

AC AAT06947;

DT 26-JUN-1996 (first entry)

DE C-promoter binding factor 1 coding sequence.

XX C-promoter binding factor 1; CBP1; transcription factor; TF; EBNA2;

XX Epstein-Barr virus; CBP1-TF transcription complex; viral infection;

XX bacterial infection; fungal infection; metabolic disease; inflammation;

XX genetic disease; cell growth dysfunction; regulatory dysfunction;

XX neoplasm; hypersensitivity; human; animal; plant; therapy; ds.

XX Synthetic.

XX MO9532307-A1.

XX 30-NOV-1995.

XX 17-MAY-1995; 95WO-US05966.

XX 20-MAY-1994; 94US-0246977.

XX (TULAR-) TULARIK INC.

XX Henkel T, Peterson MG;

XX WPI; 1996-020599/02.

XX P-PSDB; AAR86790.

XX Screening cpds. which disrupt complex formation between C-promoter

XX binding factor and transcription factor - potentially useful for

XX treating diseases involving the expression of a gene modulated by

XX such complex formation

XX Claim 1; Page 22-24; 34pp; English.

XX This sequence represents the cDNA encoding C-promoter binding factor-1

XX (CBP1). A partially purified form of the protein encoded by this screening

XX is used in the method of the invention. The method is used for screening

XX a chemical library for pharmacological agents. The protein is able to

DB 543 CTCATGTT 551

RESULT 6
AAK40082

ID AAK40082 standard; DNA; 666 BP.

XX AAK40082;

XX 02-JUL-1999 (first entry)

XX Gastric cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

XX Homo sapiens.

XX WO9904265-A2.

XX 28-JAN-1999.

XX 15-JUL-1998; 98WO-US14679.

XX 22-JUN-1998; 98US-0102322.

XX 17-JUL-1997; 97US-0896164.

XX 10-OCT-1997; 97US-0061599.

XX 10-OCT-1997; 97US-0061765.

XX 10-OCT-1997; 97US-0948705.

XX 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,

XX Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;

XX Tureci O;

XX WPI: 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

XX isolated using sera from cancer patients, used to develop products

XX for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 691; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised

XX by expression of a human cancer associated antigen precursor coded for by

XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a

XX biological sample isolated from a subject with an agent that specifically

XX binds to the NAM, an expression product or a fragment of an expression

XX product complexed with an HLA molecule; and (b) determining the

XX interaction between the agent and the NAM or the expression product as a

XX determination of the disorder. The products and methods can be used in

XX the diagnosis, monitoring, research, or treatment of conditions

XX characterised by the expression of various cancer associated antigens.

XX The invention provides nucleic acid sequences and encoded polypeptides

XX which are cancer associated antigen precursors expressed in human breast

XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

XX lung cancer.

XX Sequence 666 BP; 214 A; 120 C; 155 G; 173 T; 4 other;

Query Match 72.5%; Score 362.6; DB 20; Length 666;

Best Local Similarity 98.9%; Pred. No. 1.8e-103;

Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 133 GCCTGGAAATTTGGTGGAGGCGCTCCACCTAAGCACTTCTAGGAACTGTGGAAA 191

DB 31 GACAGGGAATTTGGTGGAGGCGCTCCACCTAAGCACTTCTAGGAACTGTGGAAA 90

OY 192 TATTTTAAAGGAGGAGGATCAACACTACTTCTTATTCAGAAAATTCACAGAA 251

DB 91 TTTATTTAAAGAGCGAGGAGTCAACAGTACTTATTTCTTCATGCAAAAAGTTGCACAGAA 150

OY 252 GTCATATGAAATGAAATGAAATGTTTGTGGCCACCTCTCTGTGTATCTTATGAGGAG 311

DB 151 GTCATATGAAATGAAATGAAATGTTTGTGGCCACCTCTCTGTGTATCTTATGAGGAG 210

OY 312 CGGATGGAAGAAAAGAAAAGAAAAGTGAACGCGATGCTGTGTTCAGAACAGATCTCA 371

DB 211 TGGATGGAAGAAAAGAAAAGAAAAGTGAACGCGATGCTGTGTTCAGAACAGATCTCA 270

OY 372 ACCGTGTGCAATTTATGGAATAGTACCAAGAAATGCACACCTTAATTTGGA 431

DB 271 ACCGTGTGCAATTTATGGAATAGTACCAAGAAATGCACACCTTAATTTGGA 330

OY 432 AGGAAAGCACTATTGCAAGCCAAACATTTGTATATCTGACTGAGACAGAAAGCA 491

DB 331 AGGAAAGCACTATTGCAAGCCAAACATTTGTATATCTGACTGAGACAGAAAGCA 390

OY 492 CTCATTTT 500

DB 391 CTCATGTT 399

RESULT 7

AAK40076

ID AAK40076 standard; DNA; 707 BP.

XX AAK40076;

XX 02-JUL-1999 (first entry)

XX Gastric cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

XX Homo sapiens.

XX WO9904265-A2.

XX 28-JAN-1999.

XX 15-JUL-1998; 98WO-US14679.

XX 22-JUN-1998; 98US-0102322.

XX 17-JUL-1997; 97US-0896164.

XX 10-OCT-1997; 97US-0061599.

XX 10-OCT-1997; 97US-0061765.

XX 10-OCT-1997; 97US-0948705.

XX 11-OCT-1997; 97GB-0021697.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

XX Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;

XX Tureci O;

XX WPI: 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

XX isolated using sera from cancer patients, used to develop products

XX for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 689-689; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised

XX by expression of a human cancer associated antigen precursor coded for by

XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a

XX biological sample isolated from a subject with an agent that specifically

XX binds to the NAM, an expression product or a fragment of an expression

XX product complexed with an HLA molecule; and (b) determining the

XX interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterized by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX Sequence 707 BP; 227 A; 129 C; 155 G; 180 T; 16 other;

Query Match 70.9%; Score 354.4; DB 20; Length 707;
 Best Local Similarity 97.0%; Pred. No. 7.1e-101;

Matches 358; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 132 GCGTGGGAATTTGGTGGAGCGGCTCCACCTAAGCACTTACGAGGAGGAGGAGGAA 191
 DB 17 GACAGGAAATTTGGTGGAGCGGCTCCACCTAAGCACTTACGAGGAGGAGGAGGAA 76

QY 192 TTATTTAAAGAGCGGAGGATCAACAGTCTTATCTTATGCAAAAGTGCACAGAA 251
 DB 77 TTATTTAAAGAGCGGAGGATCAACAGTCTTATCTTATGCAAAAGTGCACAGAA 136

QY 252 GTCATATGGAATGAAAAAGTTTTCGCCACCTCTTGTGTATATCTTATGAGCGAG 311
 DB 137 GTCATATGGAATGAAAAAGTTTTCGCCACCTCTTGTGTATATCTTATGAGCGAG 196

QY 312 CGATGGAAGAAAAAAGAACAAATGAGAGCGATGTTGTGTGACAGAGAGCTCA 371
 DB 197 TGGATGGAAGAAAAAAGAACAAATGAGAGCGATGTTGTGTGACAGAGAGCTCA 256

QY 372 ACCGTGCACTTATGAGATGAGAAATGAGACCAAGAAATGACAGCTTAACCTGGA 431
 DB 257 ACCGTGCACTTATGAGATGAGAAATGAGACCAAGAAATGACAGCTTAACCTGGA 316

QY 432 AGGAAAGACTATGACAGCCAAACATTTATATCTGATCTGACAGACGAGGAGCA 491
 DB 317 AGGAAAGACTATGACAGCCAAACATTTATATCTGATCTGACAGACGAGGAGCA 376

QY 492 CTTCATTTT 500
 DB 377 CTTCATTTT 385

RESULT 8

AAE22790 AAF22790 standard; cDNA; 1336 BP.

AC AAF22790;

DE 26-MAR-2001 (first entry)

DE Human prostate cancer associated antigen nucleotide sequence SEQ ID:369.

KW Human: breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytostatic; cancer vaccine; ss.

OS Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000MO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Obata Y;

DR WPI; 2001-025274/03.

PT Nucleic acids encoding breast, gastric and prostate cancer associated
 PT antigen precursors, useful for diagnosing and treating a condition
 PT characterized by expression of an abnormal amount of a protein, e.g.
 PT cancer -

XX Claim 50; Page 395; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

CC represent nucleotide sequences encoding human breast, gastric and

CC prostate cancer associated antigen precursors (CAAP), respectively.

CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

CC represent human breast, gastric and prostate CAAP protein sequence

CC respectively. CAAPs have cytostatic activity and can be used in the

CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterized by expression of an abnormal amount of a protein,

XX e.g. cancer.

XX Sequence 1336 BP; 400 A; 239 C; 356 G; 333 T; 8 other;

Query Match 69.8%; Score 348.8; DB 22; Length 1336;

Best Local Similarity 99.2%; Pred. No. 5.6e-99;

Matches 361; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 137 GGAATTTGGTGGAGCGGCTCCACCTAAGCACTTACGAGGAGGAGGAGGAA 196

DB 412 GGAATTTGGTGGAGCGGCTCCACCTAAGCACTTACGAGGAGGAGGAGGAA 471

QY 197 TAAAGAGCGGAGGATCAACAGTCTTATCTTATGCAAAAGTGCACAGAGCTAT 256

DB 472 TAAAGAGCGGAGGATCAACAGTCTTATCTTATGCAAAAGTGCACAGAGCTAT 531

QY 257 ATGGAATGGAATGAAAAAGTTTTCGCCACCTCTTGTGTATATCTTATGAGCGAG 316

DB 532 ATGGAATGGAATGAAAAAGTTTTCGCCACCTCTTGTGTATATCTTATGAGCGAG 591

QY 317 GGAAGAAAAAAGAACAAATGAGAGCGATGTTGTGTGACAGAGAGCTCAACCGT 376

DB 592 GGAAG-AAAAAAGAACAAATGAGAGCGATGTTGTGTGACAGAGAGCTCAACCGT 650

QY 377 GTGCATTTATGGAATGAGAAATGAGACCAAGAAATGACAGCTTAACCTGGAAGAA 436

DB 651 GTGCATTTATGGAATGAGAAATGAGACCAAGAAATGACAGCTTAACCTGGAAGAA 710

QY 437 AGAATCTATGACAGCCAAACATTTATATCTGATCTGACAGACGAGGAGCACTTCA 496

DB 711 AGAATCTATGACAGCCAAACATTTATATCTGATCTGACAGACGAGGAGCACTTCA 770

QY 497 TTTT 500

DB 771 TGT 774

RESULT 9

AAH42783 AAF42783 standard; DNA; 373 BP.

AC AAF42783;

DE 01-OCT-2001 (first entry)

DE Nucleotide sequence of JK-recombination signal binding protein gene.

KW Cancer associated gene; prostate cancer; stomach cancer;

KW oesophageal cancer; cancer; vaccine; ss.

OS Homo sapiens.

PN WO200153524-A2.

PD 26-JUL-2001.

PF 18-JAN-2001; 2001MO-GB00188.

XX 18-JAN-2000: 2000GB-0000993.
 PR (UNO-) UNIV NOTTINGHAM TRENT.
 PA
 XX Rees RC, LI G, Mian S;
 PI
 XX WPI: 2001-476121/51.
 DR
 XX Use of cancer-associated genes and their products in detecting,
 PT monitoring, treating or preventing cancer, specifically prostate
 PT cancer, and in developing DNA-based vaccines that promote anti-tumor
 PT responses
 PS
 XX Claim 1; Page 18; 43pp; English.
 CC AAH42781-AAH42846 represent the nucleotide sequences of cancer
 CC associated genes, identified using SEREX (Serological Identification of
 CC Antigens by Recombinant Expression Cloning). The genes are overexpressed
 CC in prostate cancer, and some are overexpressed in other cancers such as
 CC stomach cancer and oesophageal cancer. The nucleic acids are useful for
 CC detecting and monitoring cancer, particularly prostate cancer. They are
 CC particularly useful in the treatment or prevention of cancer, in
 CC producing DNA-based vaccines against prostate cancer or that promote
 CC anti-tumor immune responses, and to raise antibodies. The expression of
 CC genes and detection of their protein products and/or peptides may be
 CC used to monitor disease progression during therapy, or as a prognostic
 CC indicator of initial disease status of the patient.
 XX
 SO Sequence 373 BP; 128 A; 59 C; 93 G; 93 T; 0 other;
 Query Match 65.4%; Score 327.2; DB 22; Length 373;
 Best Local Similarity 98.8%; Pred. No. 1.7e-97;
 Matches 340; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 132 GCGTGGGAAATTTGGTGGGCGGCTCCACCTTAACGACTTACTAGGAGACTTATGCGGAA 191
 DB 30 GACAGGGAATTTGGTGGGCGGCTCCACCTTAACGACTTACTAGGAGACTTATGCGGAA 89
 QY 192 TTAATTTAAAGACGAGGAGATCAACAGTACTTATCTTCATGCAAAAGTTGCGACAGAA 251
 DB 90 TTAATTTAAAGACGAGGAGATCAACAGTACTTATCTTCATGCAAAAGTTGCGACAGAA 149
 QY 252 GTCAATATGGAATGAAAGGTTTGGCCACCTCTGTGTATATCTTATGCGGAG 311
 DB 150 GTCAATATGGAATGAAAGGTTTGGCCACCTCTGTGTATATCTTATGCGGAG 209
 QY 312 CGGATGGAAGAAAAAAGAACAAATGGAACGGAGTGTCTGCAACAGCTCA 371
 DB 210 TCGATGGAAGAAAAAAGAACAAATGGAACGGAGTGTCTGCAACAGCTCA 269
 QY 372 ACCGTGCAATTTTGGGATAGGAAATAGTACCAAGAAATGACAGCTTAACCTTGA 431
 DB 270 ACCGTGCAATTTTGGGATAGGAAATAGTACCAAGAAATGACAGCTTAACCTTGA 329
 QY 432 AGGAAGAACTATGTCAGACCCAAAACATTTATATATCTGACT 475
 DB 330 AGGAAG-ACATATGTCAGACCCAAAACATTTATATCTGACT 372
 RESULT 10
 AAX39641
 ID AAX39641 standard; DNA; 1599 BP.
 AC AAX39641;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Breast cancer associated gene.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.

XX Homo sapiens.
 OS
 XX MO9904265-A2.
 PN
 XX 28-JAN-1999.
 PD
 XX 15-JUL-1998; 98MO-US14679.
 PE
 XX 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX
 XX Chen Y, Gout I, Gure A, O'Hare M, Obara Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX
 XX WPI: 1999-132448/11.
 DR
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 PT
 XX
 PS Claim 67; Page 401-402; 787pp; English.
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample, isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product of a fragment of an expression
 CC product complexed with an HLA molecule, and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SO Sequence 1599 BP; 498 A; 317 C; 358 G; 426 T; 0 other;
 Query Match 64.2%; Score 320.8; DB 20; Length 1599;
 Best Local Similarity 92.6%; Pred. No. 3.7e-90;
 Matches 337; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 137 GGAATTTGGTGGAGCGGCTCCACCTTAACGACTTACTAGGAGACTTATGCGGAAATAT 196
 DB 137 GGAATTTGGTGGAGCGGCTCCACCTTAACGACTTACTAGGAGACTTATGCGGAAATAT 196
 QY 197 TAAAGAGCGAGGAGATCAACAGTACTTATCTTCATGCAAAAGTTGCGACAGAGTCA 256
 DB 197 TAAAGAGCGAGGAGATCAACAGTACTTATCTTCATGCAAAAGTTGCGACAGAGTCA 256
 QY 257 ATGGAATGGAAGAAAGGTTTGGCCACCTCTGTGTATATCTTATGCGGAGCGGAT 316
 DB 257 ATGGAATGGAAGAAAGGTTTGGCCACCTCTGTGTATATCTTATGCGGAGCGGAT 316
 QY 317 GGAAGAAAAAAGAACAAATGGAACGGAGTGTCTGCAACAGCTCAACCG 376
 DB 317 GGAAGAAAAAAGAACAAATGGAACGGAGTGTCTGCAACAGCTCAACCG 376
 QY 377 GTGCAATTTATGGATAGGAAATAGTACCAAGAAATGACAGCTTAACCTTGAAGGAA 436
 DB 377 GTGCAATTTATGGATAGGAAATAGTACCAAGAAATGACAGCTTAACCTTGAAGGAA 436
 QY 437 AGAATATTTGCGACCCAAAACATTTATATCTGACTTCAAGACAGGAGGAAAGCACTTCA 496
 DB 437 AGAATATTTGCGACCCAAAACATTTATATCTGACTTCAAGACAGGAGGAAAGCACTTCA 496

Matches 340; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

```
QY 137 GGAATTTTGGTGGGCGCTCACCCTAAAGACCTTACTAGGAGGATGCGCAATTTAT 196
    |||
Db 211 GGAATTTTGGTGGGCGCTCACCCTAAAGACCTTACTAGGAGGATGCGCAATTTAT 270
QY 197 TAAAGAGCGAGGAGTCAACACTACTTATCTTCATGCAAAAGTTCACAGAACTCAT 256
    |||
Db 271 TAAAGAGCGAGGAGTCAACACTACTTATCTTCATGCAAAAGTTCACAGAACTCAT 330
QY 257 ATGGAATGAAAAAGTTTGGCCACCTCTGTGTATATCTTATGCGACGCGAT 316
    |||
Db 331 ATGGAATGAAAAAGTTTGGCCACCTCTGTGTATATCTTATGCGACGCGAT 390
QY 317 GGAAGAAAAAAGAACATGGAACGCGATGTTGTTCTGAAACAGATCTCAACCGT 376
    |||
Db 391 GGAAG-AAAAAAGAACATGGAACGCGATGTTGTTCTGAAACAGATCTCAACCGT 449
QY 377 GTGCATTTATTTGGTATGGAATGTAACCAAGAAAGCAAGCTTAACCTGGAAGGA 436
    |||
Db 450 GTGCATTTATTTGGTATGGAATGTAACCAAGAAAGCAAGCTTAACCTGGAAGGA 509
QY 437 AGAATATTGACAGCCAAACATTGTATATCTGACTGACAGCAAGCAAA 488
    |||
Db 510 AGAATATTGACACA-CCAAACATTGTATATCTGACTGACAGCAAGCAAGCAAA 560
```

RESULT 13

AAF22795 standard; cDNA: 768 BP.

AAF22795;

26-MAR-2001 (first entry)

Human prostate cancer associated antigen nucleotide sequence SFO ID:374.

Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytosolic; cancer vaccine; ss.

OS Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Obata Y.

PI MPI: 2001-025274/03.

PT Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition

PT characterized by expression of an abnormal amount of a protein, e.g.

PS cancer -

PS Claim 50; Page 396-397; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

XX CC represent nucleotide sequences encoding human breast, gastric and

XX CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

XX CC represent human breast, gastric and prostate CAAP protein sequence

XX CC respectively. CAAPs have cytosolic activity and can be used in the

XX CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC e.g. cancer.

SO Sequence 768 BP; 254 A; 139 C; 169 G; 204 T; 2 other;

Query Match 56.6%; Score 282.8; DB 22; Length 768;

Best local similarity 99.3%; Pred. No. 2, 1e-78;

Matches 284; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 215 AAACAGTACTTATTTCTTCATGCAAAAGTGCACAGAGTCAATATGCAAAAGT 274

Db 1 AAACAGTACTTATTTCTTCATGCAAAAGTGCACAGAGTCAATATGCAAAAGT 60

QY 275 TTTTGGCCACCTCTGTGTATATCTTATGGAACGCGATGGAAGAAAAAAGAAC 334

Db 61 TTTTGGCCACCTCTGTGTATATCTTATGGAACGCGATGGAAGAAAAAAGAAC 120

QY 335 AAATGGAACGCGATGTTGTTCTGAAACAGATCTCAACCGTGTGATTTATGGAATG 394

Db 121 AAATGGAACGCGATGTTGTTCTGAAACAGATCTCAACCGTGTGATTTATGGAATG 180

QY 395 GAAATATGACCAAGAAATGCAAGCTTAACCTTGAAGGAAAGAACTTATGCAAGCA 454

Db 181 GAAATATGACCAAGAAATGCAAGCTTAACCTTGAAGGAAAGAACTTATGCAAGCA 240

QY 455 AAACATTTATATCTGACTGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 500

Db 241 AAACATTTATATCTGACTGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 286

RESULT 14

AAF22792 standard; cDNA: 625 BP.

AAF22792;

26-MAR-2001 (first entry)

Human prostate cancer associated antigen nucleotide sequence SFO ID:371.

Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytosolic; cancer vaccine; ss.

OS Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Obata Y.

PI MPI: 2001-025274/03.

PT Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition

PT characterized by expression of an abnormal amount of a protein, e.g.

PS cancer -

PS Claim 50; Page 395-396; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

XX CC represent nucleotide sequences encoding human breast, gastric and

XX CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

XX CC represent human breast, gastric and prostate CAAP protein sequence

XX CC respectively. CAAPs have cytosolic activity and can be used in the

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
e.g. cancer.

SQ Sequence 625 BP; 205 A; 103 C; 142 G; 164 T; 11 other;

Query Match 56.4%; Score 281.8; DB 22; Length 625;
Best Local Similarity 91.8%; Pred No. 3.9e-78;
Matches 315; Conservative 0; Mismatches 25; Indels 3; Gaps 2;

QY 137 GGAATTTGGTGGAGCGCCCTCACTAAAGACTTACAGGAGAGCTATGCGAAATATT 196
DB 208 GGAATTTGGTGGAGCGCCCTCACTAAAGACTTACAGGAGAGCTATGCGAAATATT 267
QY 197 TAAAGAGCGAGGGATCAACAGTACTTATCTTCATGCAAAATGTCACAGAAATCAT 256
DB 268 TAAAGAGCGAGGGATCAACAGTACTTATCTTCATGCAAAATGTCACAGAAATCAT 327
QY 257 ATGGAATGAAAAAGTTTTTTTGGCCACCTCTGTGTATATCTATGCGAGCGGAT 316
DB 328 ATGGAATGAAAAAGTTTTTTTGGCCACCTCTGTGTATATCTATGCGAGGAT 387
QY 317 GGAAGAAAAAAGAACAAATGGAACGCGATGTTGTTGACAGAGA--GTCTCAAC 374
DB 388 GGAAGAAAAAAGAACAAATGGAACGCGATGTTGTTGACAGAGA--GTCTCAAC 447
QY 375 GTGTGATTTTATTTGGATAGGAATAGTGACCAAGAAATGACAGAGCTAACTTGAAG 434
DB 448 GATGCTATTTATTTGGATAGGAATAGTGACCAAGAAATGACAGAGCTAACTTGAAG 506
QY 435 AAGAACTATTGTCAGCCCAAAACATTTGATATATCTACTCA 477
DB 507 AAGAACTATTGTCAGCCCAAAACATTTGATATATCTACTCA 549

RESULT 15

AAF22791 ID AAF22791 standard; cDNA; 815 BP.

XX AAF22791;

XX AC 26-MAR-2001 (first entry)

XX DE Human prostate cancer associated antigen nucleotide sequence SEQ ID:370.
XX DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX OS Homo sapiens.

XX PN W0200073801-A2.

XX PD 07-DEC-2000.

XX PF 26-MAY-2000; 2000MO-US14749.

XX PR 28-MAY-1999; 99US-0136526.

XX PR 10-SEP-1999; 99US-0153454.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Obata Y;

XX DR WPI; 2001-025274/03.

XX PT Nucleic acids encoding breast, gastric and prostate cancer associated
XX PT antigen precursors, useful for diagnosing and treating a condition
XX PT characterized by expression of an abnormal amount of a protein, e.g.
XX PT cancer -

XX PS Claim 50: Page 395; 799pp; English.

XX CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.

SQ Sequence 815 BP; 261 A; 149 C; 179 G; 218 T; 8 other;

Query Match 47.8%; Score 238.8; DB 22; Length 815;
Best Local Similarity 98.8%; Pred. No. 1.4e-64;
Matches 251; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 247 CAGAGTCATATGGAATGAAAAAGTTTTTTGGCCACCTCTGTGTATATCTTATG 306
DB 1 CAGAGTCATATGGAATGAAAAAGTTTTTTGGCCACCTCTGTGTATATCTTATG 60
QY 307 GCGAGCGATGGAAGAAAAAAGAACAAATGGAACGCGATGTTGTTGACAGAG 366
DB 61 GCGAGCGATGGAAG--AAAAAAGAACAAATGGAACGCGATGTTGTTGACAGAG 119
QY 367 TCTCAACCGTGTGATTTATTTGGATAGGAATAGTGACCAAGAAATGACAGAGCTAAAC 426
DB 120 TCTCAACCGTGTGATTTATTTGGATAGGAATAGTGACCAAGAAATGACAGAGCTAAAC 179
QY 427 TTGGAAGAAAGAACTATGTCAGCCCAAAACATTTGATATATCTACTCAAGAGCA 486
DB 180 TTGGAAGAAAGAACTATGTCAGCCCAAAACATTTGATATATCTACTCAAGAGCA 239
QY 487 AAGCACTTCATTTT 500
DB 240 AAGCACTTCATGTT 253

Search completed: October 10, 2002, 15:15:40
Job time : 90 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 18.7333 Seconds

(without alignments)
6556.054 Million cell updates/sec

Title: US-09-489-101a-9_COPY_1_500

Perfect score: 500
Sequence: 1 gaattcctgactgacacagc.....cgatatgattcgtatg 500Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	29.8	6.0	1541	US-08-200-232-3	Sequence 3, Appl
C 2	29.8	6.0	1541	PCT-US95-02219-3	Sequence 3, Appl
C 3	29.8	6.0	1541	PCT-US95-02219A-3	Sequence 1, Appl
C 4	29.8	6.0	3590	US-08-587-889-1	Sequence 1, Appl
C 5	29.8	6.0	3590	PCT-US96-09193-1	Sequence 6, Appl
C 6	29.6	5.9	4495	US-09-417-822-7	Sequence 7, Appl
C 7	29.6	5.9	4534	US-09-417-822-6	Sequence 8, Appl
C 8	29.6	5.9	4535	US-09-417-822-8	Sequence 8, Appl
C 9	29.6	5.9	5117	PCT-US95-05512-1	Sequence 1, Appl
C 10	29.6	5.9	5117	US-09-417-822-1	Sequence 1, Appl
C 11	29.6	5.9	663	US-08-463-115-89	Sequence 89, Appl
C 12	29.4	5.9	663	US-08-463-115-88	Sequence 88, Appl
C 13	29.4	5.9	669	US-08-463-115-88	Sequence 88, Appl
C 14	29.4	5.9	669	US-08-463-115-88	Sequence 88, Appl
C 15	29.4	5.9	669	US-08-463-115-88	Sequence 88, Appl
C 16	28.6	5.7	864	US-08-375-709-8	Sequence 8, Appl
C 17	28.6	5.7	37895	US-08-375-709-8	Sequence 8, Appl
C 18	28.6	5.7	37895	US-08-375-709-1	Sequence 1, Appl
C 19	28.6	5.7	37895	US-08-375-709-1	Sequence 1, Appl
C 20	28.6	5.7	37895	US-09-090-793-1	Sequence 1, Appl
C 21	28.4	5.7	1554	US-08-370-975B-10	Sequence 10, Appl
C 22	28.4	5.7	3494	US-09-139-802-200	Sequence 200, App
C 23	28.4	5.7	26764	US-08-370-975B-1	Sequence 1, Appl
C 24	28.4	5.7	3226	US-08-870-126-10	Sequence 10, Appl
C 25	27.8	5.6	2020	PCT-US95-13975-1	Sequence 1, Appl
C 26	27.8	5.6	6176	US-08-911-321-6	Sequence 6, Appl
C 27	27.6	5.5	2232	US-08-196-989B-3	Sequence 3, Appl

C 28	27.6	5.5	2232	US-08-760-936-3	Sequence 3, Appl
C 29	27.6	5.5	3516	US-09-188-930-257	Sequence 257, Appl
C 30	27.6	5.5	3706	US-08-207-904-16	Sequence 16, Appl
C 31	27.6	5.5	7305	US-08-286-740-4	Sequence 4, Appl
C 32	27.6	5.5	7305	PCT-US95-09576-4	Sequence 4, Appl
C 33	27.4	5.5	1673	US-08-294-143-4	Sequence 4, Appl
C 34	27.4	5.5	1673	US-08-294-143-4	Sequence 4, Appl
C 35	27.4	5.5	1673	US-08-294-143-4	Sequence 4, Appl
C 36	27.4	5.5	4560	US-08-256-331-4	Sequence 4, Appl
C 37	27.4	5.5	5635	US-08-136-742A-3	Sequence 3, Appl
C 38	27.4	5.5	5635	US-09-248-026-3	Sequence 3, Appl
C 39	27.4	5.5	5635	PCT-US93-11667-3	Sequence 3, Appl
C 40	27.4	5.5	6126	US-08-951-912-3	Sequence 3, Appl
C 41	27.4	5.5	6126	US-09-174-077-3	Sequence 3, Appl
C 42	27.4	5.5	6129	US-07-637-621-1	Sequence 1, Appl
C 43	27.4	5.5	6129	US-08-136-742A-1	Sequence 1, Appl
C 44	27.4	5.5	6129	US-08-135-809A-1	Sequence 1, Appl
C 45	27.4	5.5	6129	US-08-951-912-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-200-232-3/C
Sequence 3, Application US/08200232
Patent No. 5721349
GENERAL INFORMATION:
APPLICANT: Cover, Timothy L.
TITLE OF INVENTION: VACCINATING TOXIN-DEFICIENT H. PYLORI
TITLE OF INVENTION: AND RELATED METHODS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,232
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1541 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1541
US-08-200-232-3
Query Match 6.0%; Score 29.8; DB 1; Length 1541;
Best Local Similarity 56.7%; Pred.No. 5.2;
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 168 TGAATTGATGCCAAGCGTTGGTGGCGCTGACTTCTTCTGATGGGATGTCAT 227

Db 1399 TGTGTGTTATTAAGGGGCTAGGGGCTCTTTGATTTGCTCTTGCAGATTGGTGTGCTTG 1340

OY 228 CTCGAACATCAATTATTCACACAGAGCCCTTAGGAAAC 264

Db 1339 CAGAAATATTGTCAATACCTAGCGGCTTGGACATCCGAC 1303

RESULT 2
 PCT-US95-02219-3/C
 Sequence 3, Application PC/TUS9502219
 GENERAL INFORMATION:
 APPLICANT: Cover, Timothy L.
 APPLICANT: Blaser, Martin J.
 TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI
 TITLE OF INVENTION: AND RELATED METHODS
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG P.C.
 STREET: 127 Peachtree Street, Suite 1200
 City: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/02219
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Spratt, Gwendolyn D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.023
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1541 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1541
 PCT-US95-02219-3

Query Match	Similarity	6.0%;	Score 29.8;	DB 5;	Length 1541;
Best Local	Similarity	56.7%;	Pred. No. 5.2;		
Matches	55;	Conservative	0;	Mismatches	42;
				Indels	0;
				Gaps	0;
QY	168	TCATGTGATGCCACGCTTGGTGGCGCTCGACCTTCCTGCTGATGAGGATGCAT	227		
Db	1399	TCGTGTGTTATTAAGGGCTAGCGCGCTCTTGTGAATGCTCTTCGACATTTGGTGTGCTG	1340		
QY	228	CTCAACATCAATATTCACGACGAGCCCTCGAGSAAAC	264		
Db	1339	CAGAAATATTGTCTACTAATGCTAGCGCCTTGCACTCCAC	1303		

RESULT 3

PCT-US95-02219A-3/c

Sequence 3, Application PC/TU9502219A

GENERAL INFORMATION:

APPLICANT: Cover, Timothy L.

APPLICANT: Tummuuru, Murali KR

APPLICANT: Cao, Ping

APPLICANT: Thompson, Stuart A.

1 APPLICANT: Blaser, Martin J.
2 TITLE OF INVENTION: VACCINATING TOXIN-DEFICIENT H. PYLORI
3 TITLE OF INVENTION: AND THE RELATED METHODS
4 NUMBER OF SEQUENCES: 19
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: NEEDLE & ROSENBERG P.C.
7 STREET: 127 Peachtree Street, Suite 1200
8 City: Atlanta
9 STATE: Georgia
10 COUNTRY: USA
11 ZIP: 30303
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentin Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: PCT/US95/02219A
20 FILING DATE:
21 CLASSIFICATION:
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Spratt, Gwendolyn D.
24 REGISTRATION NUMBER: 36,016
25 REFERENCE/DOCKET NUMBER: 2200.023
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 404/688-0770
28 TELEFAX: 404/688-9880
29 INFORMATION FOR SEQ ID NO: 3:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 1541 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 FEATURE:
37 NAME/KEY: CDS
38 LOCATION: 1..1541
39
40 PCT-US95-02219A-3

```

Query Marc3          5.0%; Score 29.8; DB 5; Length 1541;
Best Local Similarity 56.7%; Pred. No. 5.2;
Matches 55; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 168 TGAATGATGCCAGGCTTGATGGTGGCTGCATCTTCTGATGATGGGATGCAT 227
      || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1339 TGTGTGTGTATATAAGGGTAGGCGCCTCTTGAATGTCTGACAGATGTGTGTG 1340

OY 228 CTCACATCATATATCACCAGAGCCCTTGAGSACAC 264
      | ||| | || | || | || | || | || | || | || | || | || |
Db 1339 CAGAAATATTGTCACTAGCGCCTTGACTCCAC 1303

RESULT 4
US-08-587-889-1
? Sequence 1, Application US/08587889
? Patent No. 5654397
? GENERAL INFORMATION:
? APPLICANT: CAO, Zhaodan
? APPLICANT: GROSTON, Glenn E.
? APPLICANT: GOMDEL, David V.
? TITLE OF INVENTION: INTERLEUKIN-1 RECEPTOR-ASSOCIATED
? TITLE OF INVENTION: PROTEIN KINASE AND BINDING ASSAY
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: FLEHR, HOHBACH, TEST, ALPBITTON & HERBERT
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94111-4187
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,889
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-60916
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8770
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3590 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-587-889-1

Query Match
Best Local Similarity 55.2%; Pred. No. 7.8;
Matches 58; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 231 AATCATCATATACACAGAGCCCTTGAGAACAGAGTGTGACGAGAGCTGAAAAAAA 290
DB 1096 AGACATCAAGATTCACAGCTCTTCGTGAGAGAGCTGACACCCAGCTGGAGACTT 1155
QY 291 TGTTCGTGATGAGATGATCTTCAGAGATCATCCCCAGCCA 335
DB 1156 TGCCTGCGCCGCTTCAGCCCTTTCGCGGCTCCAGCCAGCCA 1200

RESULT 5
PCT-US96-09193-1
Sequence 1, Application PC/TUS9609193
GENERAL INFORMATION:
APPLICANT: Tularik, Inc.
TITLE OF INVENTION: INTERLEUKIN-1 RECEPTOR-ASSOCIATED
TITLE OF INVENTION: PROTEIN KINASE AND BINDING ASSAY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/09193
FILING DATE: JUNE 5 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Serial No. 08/587,889
FILING DATE: JAN 16 1996
CLASSIFICATION:
APPLICATION NUMBER: U.S. Serial No. 08/494,006
FILING DATE: JUNE 23 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Brezner
REGISTRATION NUMBER: 24,774
REFERENCE/DOCKET NUMBER: PP-62191-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3590 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US96-09193-1

Query Match
Best Local Similarity 55.2%; Pred. No. 7.8;
Matches 58; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 231 AATCATCATATACACAGAGCCCTTGAGAACAGAGTGTGACGAGAGCTGAAAAAAA 290
DB 1096 AGACATCAAGATTCACAGCTCTTCGTGAGAGAGCTGACACCCAGCTGGAGACTT 1155
QY 291 TGTTCGTGATGAGATGATCTTCAGAGATCATCCCCAGCCA 335
DB 1156 TGCCTGCGCCGCTTCAGCCCTTTCGCGGCTCCAGCCAGCCA 1200

RESULT 6
US-09-417-822-6/c
Sequence 6, Application US/09417822
Patent No. 6344549

GENERAL INFORMATION:
APPLICANT: Keegan, Kathy
TITLE OF INVENTION: ATR-2
FILE REFERENCE: 27866/35633
CURRENT APPLICATION NUMBER: US/09/417,822
CURRENT FILING DATE: 1999-10-14
NUMBER OF SEQ ID NOS: 43
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 4495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: PMW-312r.3
US-09-417-822-6

Query Match
Best Local Similarity 45.6%; Pred. No. 10;
Matches 104; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 62 AATCAAGAACATTTACTGAGTCAAGTGGCGCCCTTACATCTCTCATGTGTGTTTC 121
DB 315 AAGGAATTTGATTTCAACAGTTTCCAAATGCTTGTGATGATGAGTTCCACTGTGCACT 256
QY 122 GAATATATACATCAGAGCTCTATGATCATCTGGAGATGTCTCCGTGATGTTGATGCCA 181
DB 255 ACTTGACACGCGCAAAATGACACATCTCTGTAAAGATCATCTGGCAAGAGAGAGAC 196
QY 182 AGCGTTGTCGCTCTGACATCTTCTGTGTATGAGGATGTCTCAAAATCATATA 241
DB 195 AAGCATGTAGACGTGCTGATGCTGCGACACATATTTTACATCTTCAACGAGAGTGTG 136
QY 242 TCACGAGAGCCCTTGAGACACAGGTTGAGACGAGGAGAGTGAAGAAAAA 289
DB 135 GCTAAAGCCCTTTCAGAAATTTATGAAAGACGAGAGAGAGATATAA 88

RESULT 7
US-09-417-822-7/c
Sequence 7, Application US/09417822
Patent No. 6344549
GENERAL INFORMATION:
APPLICANT: Keegan, Kathy
TITLE OF INVENTION: ATR-2
FILE REFERENCE: 27866/35633

;; CURRENT APPLICATION NUMBER: US/09/417,822
;; CURRENT FILING DATE: 1999-10-14
;; NUMBER OF SEQ ID NOS: 43
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 7
;; LENGTH: 4534
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: pmtw-312r-5
US-09-417-822-7

Query Match 5.9%; Score 29.6; DB 4; Length 4534;
Best Local Similarity 45.6%; Pred. No. 10;
Matches 104; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 62 AATCAAGACATTTACTGAAGTCAAGTGCAGCCGCTCATCTCTCATGTGCTTC 121
DB 315 AAGGAAATGATTTCACAGTTTCCAAATGCTGTGGAATACAGATTCACAGTGCAC 256
QY 122 GAATTAATTCATCAGAGCTCTATCGATCGAGATGCTCTCCGATGTTGATGCCA 181
DB 255 AGTTGAACACGGCAACATCGACACATCTCTGTAAAGATCATCTGCGAAGAGAGGC 196
QY 182 AGGCTTTGGTGGCTGCTGCTCTCTGTGTATGGGATGTCATCTCAACATCAATA 241
DB 195 AGAGCATGTAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 136
QY 242 TCACGAGAGCCCTTGAGACAGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
DB 135 GCTAAAGAGCTTTGCAAAATTTATGAAAGACGAGAGAGAGAGAGAGAGAGAG 88

RESULT 8

US-09-417-822-8/c
;; Sequence 8, Application US/09417822
;; Patent No. 6344549
;; GENERAL INFORMATION:
;; APPLICANT: Keegan, Kathy
;; TITLE OF INVENTION: ATR-2
;; FILE REFERENCE: 27866/3563
;; CURRENT APPLICATION NUMBER: US/09/417,822
;; CURRENT FILING DATE: 1999-10-14
;; NUMBER OF SEQ ID NOS: 43
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 8
;; LENGTH: 4535
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: pmtw-312r-7
US-09-417-822-8

Query Match 5.9%; Score 29.6; DB 4; Length 4535;
Best Local Similarity 45.6%; Pred. No. 10;
Matches 104; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 62 AATCAAGACATTTACTGAAGTCAAGTGCAGCCGCTCATCTCTCATGTGCTTC 121
DB 315 AAGGAAATGATTTCACAGTTTCCAAATGCTGTGGAATACAGATTCACAGTGCAC 256
QY 122 GAATTAATTCATCAGAGCTCTATCGATCGAGATGCTCTCCGATGTTGATGCCA 181
DB 255 AGTTGAACACGGCAACATCGACACATCTCTGTAAAGATCATCTGCGAAGAGAGGC 196
QY 182 AGGCTTTGGTGGCTGCTGCTCTCTGTGTATGGGATGTCATCTCAACATCAATA 241
DB 195 AGAGCATGTAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 136
QY 242 TCACGAGAGCCCTTGAGACAGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
DB 135 GCTAAAGAGCTTTGCAAAATTTATGAAAGACGAGAGAGAGAGAGAGAGAGAG 88

RESULT 9
US-08-854-585-1/c
;; Sequence 1, Application US/08854585
;; Patent No. 6114140
;; GENERAL INFORMATION:
;; APPLICANT: Tonks, Nicholas K. and Stman, Arne
;; TITLE OF INVENTION: Density Enhanced Protein Tyrosine Phosphatase
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 233 South Wacker Drive, Suite 6300
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/854,585
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/237,940
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Borun, Michael F.
;; REGISTRATION NUMBER: 25,447
;; REFERENCE/DOCKET NUMBER: 27866/31954
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-474-0448
;; TELEFAX: 312-474-0448

INFORMATION FOR SEQ ID NO: 1

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5117 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 350..4364
US-08-854-585-1

Query Match 5.9%; Score 29.6; DB 3; Length 5117;
Best Local Similarity 57.6%; Pred. No. 11;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 85 TCAAGTGGTGGCGGCTCATCTCTCATGATGCTTGAATATATATACAGAGCTCTAT 144
DB 2295 TCAAGTGGTGGCGGCTCATCTCTCATGATGCTTGAATATATATATACAGAGCTCTAT 176
QY 145 CGATCAGTGGAGATGCTCTCTCGATGTTGA 176
DB 2235 ACATTCGTGGCGGCTGTCTACTGTGAGTGA 2204

RESULT 10

PCT-US95-05512-1/c
;; Sequence 1, Application PC/RUS9505512
;; GENERAL INFORMATION:
;; APPLICANT: Tonks, Nicholas K. and Stman, Arne
;; TITLE OF INVENTION: Density Enhanced Protein Tyrosine
;; TITLE OF INVENTION: Phosphatase
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: Marshall, O'Toole, Gerstein, Murray &
;; ADDRESS: Borun
;; STREET: 233 South Wacker Drive, Suite 6300
;; CITY: Chicago
;; STATE: Illinois

COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05512
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27866/31954
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5117 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 350..4364
PCT-US95-05512-1

Query Match 5.9%; Score 29.6; DB 5; Length 5117;
Best Local Similarity 57.6%; Pred. No. 11;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 85 TCAGAGTGTGCGCCCTACATCTCTCAATGTGTGCAATATTCATAGCTCTT 144
DB 2295 TCAGAGTGTGCGCCCTACATCTCTCTGTGTGTGCTACTACATCAATGTG 2236
OY 145 CGATCACTGCGAGATGTCTCTCGTATGTGA 176
DB 2235 ACATGTCTGGCGGTGTGTGCTGCTGCTGCA 2204

RESULT 11
US-09-417-822-1/C
Sequence 1, Application US/09417822
Patent No. 6344549
GENERAL INFORMATION:
APPLICANT: Keegan, Kathy
TITLE OF INVENTION: ATR-2
FILE REFERENCE: 27866/35633
CURRENT APPLICATION NUMBER: US/09/417,822
CURRENT FILING DATE: 1999-10-14
NUMBER OF SEQ ID NOS: 43
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 8838
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (31)..(8820)
FEATURE:
OTHER INFORMATION: ATR-2 Full Length
US-09-417-822-1

Query Match 5.9%; Score 29.6; DB 4; Length 8838;
Best Local Similarity 45.6%; Pred. No. 14;
Matches 104; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

OY 62 AAATCAAGACATTTACGAGTCAAGTGTGCGCCCTACATCTCTCAATGTGTT 121
DB 314 AAAGATTCATTTCAACAGATTTCGAATGCTGTGCAATACAGTTCACGTG 255

OY 122 GAATATTCATCAGAGCTCTATGATCAGTGGAGATGTCTCTGATGTGATGCCA 181
DB 254 AGTTGAAACGCGCAAAACATGACATCTGTAAAGATCATCTGGCAAGAGGAC 195
OY 182 AGCGTTGGTGGCGCTGATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 241
DB 194 AGAGCATGTAGACTGTCTGATCTCTGCGAGACAGATATTCATCTTCAACGAGATG 135
OY 242 TCACGAGCCCTTGAAGACAGAGTGTGAGACGGAAGCTAGAAAAA 289
DB 134 GCTTAAAGGCTTTGCAAAATTTATGAAAGACGAGAGATATAA 87

RESULT 12
US-08-463-115-89/C
Sequence 89, Application US/08463115
Patent No. 5703221
GENERAL INFORMATION:
APPLICANT: WILLIAM JOHN MARTIN
TITLE OF INVENTION: ISOLATED STEALTH VIRUSES
TITLE OF INVENTION: AND RELATED VACCINES
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 613 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,115
FILING DATE: June 5, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 08/157,811
FILING DATE: No. 5703221, member 23, 1993
APPLICATION NUMBER: 07/887,502
FILING DATE: May 22, 1992
APPLICATION NUMBER: 07/704,814
FILING DATE: May 23, 1991
APPLICATION NUMBER: 07/763,039
FILING DATE: September 20, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 213/301
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-463-115-89

Query Match 5.9%; Score 29.4; DB 1; Length 663;
Best Local Similarity 60.8%; Pred. No. 4.6;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 158 ATGCTCCGTGTGTGTGATGCCAAGCTTGGCGCTGATCTCTCTGTGTGTGT 217

Db 567 ACCTAGTGGAGATGAAGTCCACATTCTTCACAGCACTGACTCGCTGGAGATG 508
QY 218 GGGATGCACTTCACACAT 236
Db 507 AGGAACCTCTCCACAGACAT 489

RESULT 13
US-08-465-388-89/c
; Sequence 89, Application US/08465388
; Patent No. 5753488
; GENERAL INFORMATION:
; APPLICANT: WILLIAM JOHN MARTIN
; TITLE OF INVENTION: ISOLATED STEALTH VIRUSES
; TITLE OF INVENTION: AND RELATED VACCINES
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Fastseq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,388
; FILING DATE: June 5, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/157,811
; FILING DATE: No. 5753488, September 23, 1993
; APPLICATION NUMBER: 07/887,502
; FILING DATE: May 22, 1992
; APPLICATION NUMBER: 07/704,814
; FILING DATE: May 23, 1991
; APPLICATION NUMBER: 07/763,039
; FILING DATE: September 20, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 213/300
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-465-388-89

Query Match 5.9%; Score 29.4; DB 1; Length 663;
Best Local Similarity 60.8%; Pred. No. 4.6;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 158 ATGTCTCCGTGATGTGATGCCAAGGCTTGCGCTGCACTTCTTCTGTATG 217
Db 567 ACCTAGTGGAGATGAAGTCCACATTCTTCACAGCACTGACTCGCTGGAGATG 508
QY 218 GGGATGCACTTCACACAT 236
Db 507 AGGAACCTCTCCACAGACAT 489

Four

RESULT 14
US-08-463-115-88/c
; Sequence 88, Application US/08463115
; Patent No. 5703221
; GENERAL INFORMATION:
; APPLICANT: WILLIAM JOHN MARTIN
; TITLE OF INVENTION: ISOLATED STEALTH VIRUSES
; TITLE OF INVENTION: AND RELATED VACCINES
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Fastseq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,115
; FILING DATE: June 5, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/157,811
; FILING DATE: No. 5703221, September 23, 1993
; APPLICATION NUMBER: 07/887,502
; FILING DATE: May 22, 1992
; APPLICATION NUMBER: 07/704,814
; FILING DATE: May 23, 1991
; APPLICATION NUMBER: 07/763,039
; FILING DATE: September 20, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 213/301
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-463-115-88

Four

Query Match 5.9%; Score 29.4; DB 1; Length 669;
Best Local Similarity 60.8%; Pred. No. 4.6;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 158 ATGTCTCCGTGATGTGATGCCAAGGCTTGCGCTGCACTTCTTCTGTATG 217
Db 573 ACCTAGTGGAGATGAAGTCCACATTCTTCACAGCACTGACTCGCTGGAGATG 514
QY 218 GGGATGCACTTCACACAT 236
Db 513 AGGAACCTCTCCACAGACAT 495

RESULT 15
US-08-465-388-88/c
; Sequence 88, Application US/08465388
; Patent No. 5753488
; GENERAL INFORMATION:

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds
(without alignments)
9699.805 Million cell updates/sec

Title: us-09-489-101a-9_COPY_1_500

Perfect score: 500
Sequence: 1 gaattcctgactgcacagc.....cgatattgattcgtatgacg 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:
1: em_estha:
2: em_esthu:
3: em_estin:
4: em_estlu:
5: em_estlu:
6: em_estpl:
7: em_estro:
8: em_hic:
9: gb_estl:
10: gb_estl:
11: gb_hic:
12: gb_ges:
13: em_ges_hum:
14: em_ges_inv:
15: em_ges_pln:
16: em_ges_vrt:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	713	10	BE384912 601276436
2	500	100.0	872	9	AL549103 AL549103
3	500	100.0	928	9	AL556496 AL556496
4	500	100.0	1057	9	AL549090 AL549090
5	499.6	99.9	865	9	AL527428 AL527428
6	499.2	99.8	976	9	AL532402 AL532402
7	498.4	99.7	1138	10	BM464486 BM464486
8	493.2	98.6	920	9	AL518699 AL518699
9	489	97.6	837	10	BE478628 BE478628
10	488	97.6	743	10	BE408507 BE408507
11	487.4	97.5	700	10	AL553746 AL553746
12	486.4	97.3	690	10	BP976694 BP976694
13	486.4	97.3	887	10	BC685786 BC685786
14	485	97.0	936	10	BT753143 BT753143
15	484.8	97.0	955	10	BE740422 BE740422
16	483.8	96.8	1014	10	BE740422 BE740422
17					BE740422 BE740422

18	477	95.4	876	10	BE386238
19	476.4	95.3	699	10	BM008455
20	476	95.2	712	10	BF307478
21	476	95.2	798	10	BE727120
22	474.4	94.9	724	10	BE478857
23	473.8	94.8	886	10	BE409574
24	469.8	94.0	799	10	BE429542
25	462.6	92.5	870	10	BE182454
26	462	92.4	999	10	BM464540
27	460.4	92.1	752	10	BM083375
28	454	90.8	860	10	BE282087
29	453.2	90.6	797	10	BE902069
30	440.2	88.0	757	10	BE337550
31	438	87.6	859	10	BE325760
32	437.2	87.4	805	10	BE15089
33	436.8	87.4	679	10	BE620903
34	429	85.8	813	10	BE340922
35	423.8	84.8	698	10	BE409885
36	420	84.0	735	10	BE602244
37	414.6	82.9	862	10	BE475925
38	405.2	81.0	930	10	BE600224
39	397.2	79.4	722	10	BE874220
40	395.8	79.2	703	10	BE158553
41	395	79.0	892	10	BE312821
42	391	78.2	661	10	BE314366
43	386.4	77.3	728	10	BE193233
44	382.8	76.6	866	10	BE310996
45	379.4	75.9	1058	10	BE420072

ALIGNMENTS

RESULT 1
BE384912 713 bp mRNA linear EST 21-JUL-2000
LOCUS 601276436F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617337 5',
DEFINITION mRNA sequence.
ACCESSION BE384912
VERSION BE384912.1 GI:9330277
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC/DC/DMP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM284 row: h column: 10
High quality sequence stop: 611.
Location/Qualifiers
1. 713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3617337"
/clone_id="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAC(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald W. Rubin (University of

FEATURES

source

California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 186 a 162 c 170 g 195 t

ORIGIN

Query Match 100.0%; Score 500; DB 10; Length 713;
Best Local Similarity 100.0%; Pred. No. 2.3e-144;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAATTCCTGACGCGCCAGAGTGTACAGAAACATTTGCTTTGCTGGAAAGCTGCT 60
|||||
32 GAATTCCTGACGCGCCAGAGTGTACAGAAACATTTGCTTTGCTGGAAAGCTGCT 91
Db
61 CAATCAAGAACATTTACTGAAATCAAGTGGCCGCCCTACATCTCTCAATGTGCT 120
|||||
92 CAATCAAGAACATTTACTGAAATCAAGTGGCCGCCCTACATCTCTCAATGTGCT 151
Db
121 CGAATATTTACATCAAGAGCTCTATCGATCAGTGGAGATGCTCCGCGATGTTGATGCC 180
|||||
152 CGAATATTTACATCAAGAGCTCTATCGATCAGTGGAGATGCTCCGCGATGTTGATGCC 211
Db
181 AAGGCTTTGGTGGCTCTGACCTTTCTGCGTATGGGAGATGCTATCAACATCAAT 240
|||||
212 AAGGCTTTGGTGGCTCTGACCTTTCTGCGTATGGGAGATGCTATCAACATCAAT 271
Db
241 ATCAGCAGAGCCCTTGAGGAAACAGAGTTGAGACGAAAGCTAGAAAAATGTTCTGTG 300
|||||
272 ATCAGCAGAGCCCTTGAGGAAACAGAGTTGAGACGAAAGCTAGAAAAATGTTCTGTG 331
Db
301 ATCAGCAGATCTTCAAGAGATCATCCCGACGACCAACCTCGTGGCAGCAAGACAT 360
|||||
332 ATCAGCAGATCTTCAAGAGATCATCCCGACGACCAACCTCGTGGCAGCAAGACAT 391
Db
361 GTGGTAGTGGCTGTGATGATACCAAGAGGTTCTCCATTTCCAAAGACCCAGAGT 420
|||||
392 GTGGTAGTGGCTGTGATGATACCAAGAGGTTCTCCATTTCCAAAGACCCAGAGT 451
Db
421 CTCGGCGTTTGCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGTGAAGGTT 480
|||||
452 CTCGGCGTTTGCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGTGAAGGTT 511
Db
481 CGATATGATTTACTGATG 500
|||||
512 CGATATGATTTACTGATG 531
Db

RESULT 2
AL549103 872 bp mRNA linear EST 16-FEB-2001
LOCUS AL549103 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1051YK06 5
DEFINITION prime, mRNA sequence.
ACCESSION AL549103
VERSION AL549103.1 GI:12884761
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 872)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source
1..872
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1051YK06"
/clone.lib="LTI_NFL006.PL2"
/tissue_type="placenta"

BASE COUNT 203 a 212 c 236 g 216 t 5 others

ORIGIN

Query Match 100.0%; Score 500; DB 9; Length 872;
Best Local Similarity 100.0%; Pred. No. 2.5e-144;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAATTCCTGACGCGCCAGAGTGTACAGAAACATTTGCTTTGCTGGAAAGCTGCT 60
|||||
246 GAATTCCTGACGCGCCAGAGTGTACAGAAACATTTGCTTTGCTGGAAAGCTGCT 305
Db
61 CAATCAAGAACATTTACTGAAATCAAGTGGCCGCCCTACATCTCTCAATGTGCT 120
|||||
306 CAATCAAGAACATTTACTGAAATCAAGTGGCCGCCCTACATCTCTCAATGTGCT 365
Db
121 CGAATATTTACATCAAGAGCTCTATCGATCAGTGGAGATGCTCCGCGATGTTGATGCC 180
|||||
366 CGAATATTTACATCAAGAGCTCTATCGATCAGTGGAGATGCTCCGCGATGTTGATGCC 425
Db
181 AAGGCTTTGGTGGCTCTGACCTTTCTGCGTATGGGAGATGCTATCAACATCAAT 240
|||||
426 AAGGCTTTGGTGGCTCTGACCTTTCTGCGTATGGGAGATGCTATCAACATCAAT 485
Db
241 ATCAGCAGAGCCCTTGAGGAAACAGAGTTGAGACGAAAGCTAGAAAAATGTTCTGTG 300
|||||
486 ATCAGCAGAGCCCTTGAGGAAACAGAGTTGAGACGAAAGCTAGAAAAATGTTCTGTG 545
Db
301 ATCAGCAGATCTTCAAGAGATCATCCCGACGACCAACCTCGTGGCAGCAAGACAT 360
|||||
546 ATCAGCAGATCTTCAAGAGATCATCCCGACGACCAACCTCGTGGCAGCAAGACAT 605
Db
361 GTGGTAGTGGCTGTGATGATACCAAGAGGTTCTCCATTTCCAAAGACCCAGAGT 420
|||||
606 GTGGTAGTGGCTGTGATGATACCAAGAGGTTCTCCATTTCCAAAGACCCAGAGT 665
Db
421 CTCGGCGTTTGCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGTGAAGGTT 480
|||||
666 CTCGGCGTTTGCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGTGAAGGTT 725
Db
481 CGATATGATTTACTGATG 500
|||||
726 CGATATGATTTACTGATG 745
Db

RESULT 3
AL556496 928 bp mRNA linear EST 16-FEB-2001
LOCUS AL556496 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0K0061A17 5
DEFINITION prime, mRNA sequence.
ACCESSION AL556496
VERSION AL556496.1 GI:12899223
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

/note="Vector: pcwSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcwSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@life-tech.com URL :
<http://fulllength.invitrogen.com>"


```

DEFINITION   AL527428 LIT_NFL003_NBC3 Homo sapiens cDNA clone CS0DC021YU03 5
PRIME, mRNA sequence.
ACCESSION    AL527428
VERSION      AL527428.1 GI:12790921
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
AUTHORS      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC021YU03"
/clone_1ib="LIT_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com" 5 others

BASE COUNT   206 a 209 c 229 g 216 t
ORIGIN
Query Match          99.8%; Score 499.6; DB 9; Length 865;
Best Local Similarity 99.8%; Pred. No. 3.3e-144;
Matches 499; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTGACGCGCAGCTGATGACGAAACATTTGCTTTGCTGGAAGCTGCT 60
    |||||||
Db 219 GAATTCCTGACGCGCAGCTGATGACGAAACATTTGCTTTGCTGGAAGCTGCT 278

QY 61 CAATCAAGACATTTACTGAAGTCAAGTGTGCGCCCTACATCTCTCAATGTGTT 120
    |||||||
Db 279 CAATCAAGACATTTACTGAAGTCAAGTGTGCGCCCTACATCTCTCAATGTGTT 338

QY 121 CGAATTAATTAATCAAGAGCTCTATGATCATCTGAGATGTCCTCCGATGATGATGCC 180
    |||||||
Db 339 CGAATTAATTAATCAAGAGCTCTATGATCATCTGAGATGTCCTCCGATGATGATGCC 398

QY 181 AAGGCTTTGGGCTGCTGACTTTCTTCTGCTGATGAGGATGCTCAACATCAAT 240
    |||||||
Db 399 AAGGCTTTGGGCTGCTGACTTTCTTCTGCTGATGAGGATGCTCAACATCAAT 458

QY 241 ATCACCAGAGCCCTTGAGAACACAGGTTGAGACGGAACCTAGAAAAAATGTTCTGTG 300
    |||||||
Db 459 ATCACCAGAGCCCTTGAGAACACAGGTTGAGACGGAACCTAGAAAAAATGTTCTGTG 518

QY 301 ATGACGATGATCTTCAAGAGGATATGCCAGCAGCAACATCGTTGCCAGCAACAT 360
    |||||||
Db 519 ATGACGATGATCTTCAAGAGGATATGCCAGCAGCAACATCGTTGCCAGCAACAT 578

QY 361 GTGGTAGTGTCTGATAGTATACCAAAAGAGGTTCTCATTTTCAGAAAGCAGAGGT 420
    |||||||
Db 579 GTGGTAGTGTCTGATAGTATACCAAAAGAGGTTCTCATTTTCAGAAAGCAGAGGT 638

QY 421 CTCGGGAGTTTCTGATTTCTGAGCGTGTTCAGAGGAGTAGAGATGAGAGGTT 480
    |||||||

```

```

Db 639 CTCGGGAGTTTCTGATTTCTGACATTTCTCTGAGCCCTGTTTCAGGCGACTAGTATGAGAGGAGTT 698
QY 481 CGATATGATTTACTGTGATTTG 500
    |||||||
Db 699 CGATATGATTTACTGTGATTTG 718

RESULT 6
AL532402
LOCUS       AL532402 LIT_NFL001_NBC4 976 bp mRNA linear EST 13-FEB-2001
DEFINITION AL532402 LIT_NFL001_NBC4 Homo sapiens cDNA clone CS0DM014YD01 5
PRIME, mRNA sequence.
ACCESSION  AL532402
VERSION     AL532402.1 GI:12795895
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
AUTHORS    Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..976
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DM014YD01"
/clone_1ib="LIT_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com" 10 others

BASE COUNT   228 a 237 c 261 g 240 t
ORIGIN
Query Match          99.8%; Score 499.2; DB 9; Length 976;
Best Local Similarity 99.6%; Pred. No. 4.6e-144;
Matches 498; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTGACGCGCAGCTGATGACGAAACATTTGCTTTGCTGGAAGCTGCT 60
    |||||||
Db 253 GAATTCCTGACGCGCAGCTGATGACGAAACATTTGCTTTGCTGGAAGCTGCT 312

QY 61 CAATCAAGACATTTACTGAAGTCAAGTGTGCGCCCTACATCTCTCAATGTGTT 120
    |||||||
Db 313 CAATCAAGACATTTACTGAAGTCAAGTGTGCGCCCTACATCTCTCAATGTGTT 372

QY 121 CGAATTAATTAATCAAGAGCTCTATGATCATCTGAGATGTCCTCCGATGATGATGCC 180
    |||||||
Db 373 CGAATTAATTAATCAAGAGCTCTATGATCATCTGAGATGTCCTCCGATGATGATGCC 432

QY 181 AAGGCTTTGGGCTGCTGACTTTCTTCTGCTGATGAGGATGCTCAACATCAAT 240
    |||||||
Db 433 AAGGCTTTGGGCTGCTGACTTTCTTCTGCTGATGAGGATGCTCAACATCAAT 492

QY 241 ATCACCAGAGCCCTTGAGAACACAGGTTGAGACGGAAGCTAGAAAAAATGTTCTGTG 300
    |||||||
Db 493 ATCACCAGAGCCCTTGAGAACACAGGTTGAGACGGAAGCTAGAAAAAATGTTCTGTG 552

```

Query match	50.0%	Score 452.12	Length 220
Best Local Similarity	98.4%	Pred. No. 3.3e-142	
Matches 492; Conservative	5;	Mismatches 3;	Indels 0; Gaps 0

```

OY 1 GAATTCCTGACGCGCAGCTGTACAGGAACATTTGCTTTGTTGCTGGAAGCTGCT 60
    |||
Db 262 GAATTCCTGACGCGCAGCTGTACAGGAACATTTGCTTTGTTGCTGGAAGCTGCT 321
OY 61 CAAATCAAGAAACATTTACTAGTCAAGAGTGGCGCCCTACATCTCAATGTGGTT 120
    |||
Db 322 CAAATCAAGAAACATTTACTAGTCAAGAGTGGCGCCCTACATCTCAATGTGGTT 381
OY 121 CGAATTAATTACATGAGCTGTATGATGACTGGAGATGCTCTCGATGTTGATGCC 180
    |||
Db 382 CGAATTAATTACATGAGCTGTATGATGACTGGAGATGCTCTCGATGTTGATGCC 441
OY 181 AAGCTTTGGTGGCTCTGACTTTCTTCTGCTGATGATGGAGATGATCAATCAATCAAT 240
    |||
Db 442 AAGCTTTGGTGGCTCTGACTTTCTTCTGCTGATGATGGAGATGATCAATCAATCAAT 501
OY 241 ATCACCAGAGCCCTTGAGGACACAGCTTGAGAGCGAAGCAAAAAATGTTCTGTG 300
    |||
Db 502 ATCACCAGAGCCCTTGAGGACACAGCTTGAGAGCGAAGCAAAAAATGTTCTGTG 561
OY 301 ATGACGATGATCTGCAAGAGATATCCCGCAGCCCACTGCTGCGAGAGCAAT 360
    |||
Db 562 ATGACGATGATCTGCAAGAGATATCCCGCAGCCCACTGCTGCGAGAGCAAT 621
OY 361 GTGCTAGTGGCTGTGATGATACCAACAGAGGTTCTCCATTTTCAGAAAGCCAGGTT 420
    |||
Db 622 GTGCTAGTGGCTGTGATGATACCAACAGAGGTTCTCCATTTTCAGAAAGCCAGGTT 681
OY 421 CTCGCGGCTTTGCTGATTTCTGAGCTGTGTTGAGGAGAGTATGATGAGAGGTT 480
    |||
Db 682 CTCGCGGCTTTGCTGATTTCTGAGCTGTGTTGAGGAGAGTATGATGAGAGGTT 741
OY 481 CGATATGATTTACTGATG 500
    |||
Db 742 CGATATGATTTACTGATG 761

RESULT 9
LOCUS Bg478628 837 bp mRNA linear EST 21-MAR-2001
DEFINITION 60252515E1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4643907 5',
ACCESSION Bg478628
VERSION Bg478628.1 GI:13410907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 837)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgsbds-remail.nih.gov
            Tissue Procurement: ATCC/DCTD/DMP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L16M1414 row: n column: 04
            High quality sequence stop: 792.
FEATURES
    source
        1..837
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4643907"
        /clone_1db="NIH_MGC_20"
        /tissue_type="melanotic melanoma"
        /lab_host="DH10b (phage-resistant)"

```

```

/Note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using
ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      216 a      197 c      201 g      223 t
ORIGIN
Query Match      97.8%; Score 489; DB 10; Length 837;
Best Local Similarity 99.8%; Pred. No. 6,5e-141;
Matches 500; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 GAATTCCTGACTGCGACAGGTGTACAGGAACATTTGCTTTGTTGCTGGAAGCTGCT 60
    |||
Db 52 GAATTCCTGACTGCGACAGGTGTACAGGAACATTTGCTTTGTTGCTGGAAGCTGCT 111
OY 61 CAAATCAAGAAACATTTACTAGTCAAGAGTGGCGCCCTACATCTCAATGTGGTT 120
    |||
Db 112 CAAATCAAGAAACATTTACTAGTCAAGAGTGGCGCCCTACATCTCAATGTGGTT 171
OY 121 CGAATTAATTACATGAGCTGTATGATGACTGGAGATGCTCTCGATGTTGATGCC 180
    |||
Db 172 CGAATTAATTACATGAGCTGTATGATGACTGGAGATGCTCTCGATGTTGATGCC 231
OY 181 AAGCTTTGGTGGCTCTGACTTTCTTCTGCTGATGATGGAGATGATCAATCAATCAAT 240
    |||
Db 232 AAGCTTTGGTGGCTCTGACTTTCTTCTGCTGATGATGGAGATGATCAATCAATCAAT 291
OY 241 ATCACCAGAGCCCTTGAGGACACAGCTTGAGAGCGAAGCAAAAAATGTTCTGTG 300
    |||
Db 292 ATCACCAGAGCCCTTGAGGACACAGCTTGAGAGCGAAGCAAAAAATGTTCTGTG 351
OY 301 ATGACGATGATCTGCAAGAGATATCCCGCAGCCCACTGCTGCGAGAGCAAT 360
    |||
Db 352 ATGACGATGATCTGCAAGAGATATCCCGCAGCCCACTGCTGCGAGAGCAAT 411
OY 361 GTGCTAGTGGCTGTGATGATACCAACAGAGGTTCTCCATTTTCAGAAAGCCAGGTT 420
    |||
Db 412 GTGCTAGTGGCTGTGATGATACCAACAGAGGTTCTCCATTTTCAGAAAGCCAGGTT 471
OY 421 CTCGCGGCTTTGCTGATTTCTGAGCTGTGTTGAGGAGAGTATGATGAGAGGTT 480
    |||
Db 472 CTCGCGGCTTTGCTGATTTCTGAGCTGTGTTGAGGAGAGTATGATGAGAGGTT 531
OY 481 CGAT-ATGATTTACTGATG 500
    |||
Db 532 CGATATGATTTACTGATG 552

RESULT 10
LOCUS BE408507 743 bp mRNA linear EST 21-JUL-2000
DEFINITION 601303024F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637852 5',
ACCESSION BE408507
VERSION BE408507.1 GI:9344957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 743)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgsbds-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```


DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MCC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLNL at: image.lnl.gov
 Plate: LHC337 row: 0 column: 05
 High quality sequence stop: 625.
 Location/Qualifiers

FEATURES

source

1. 743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3637852"
 /clone_lib="NH_MCC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOT87; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using 2AP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 192 a 170 c 174 g 207 t
 ORIGIN

Query Match 97.6%; Score 488; DB 10; Length 743;
 Best Local Similarity 99.8%; Pred. No. 1.3e-140;
 Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 GAATTCCTGAGTCCACAGAGTACAGAAACATTTGCTTTGTCGAGAAAGTCT 60
 DB 32 GAATTCCTGAGTCCACAGAGTACAGAAACATTTGCTTTGTCGAGAAAGTCT 91
 OY 61 CAATCAAAACAACTTACTGAAGTCAAAAGTGTGCGCCCTACATCTCTCAATGTGTT 120
 DB 92 CAATCAAAACAACTTACTGAAGTCAAAAGTGTGCGCCCTACATCTCTCAATGTGTT 151
 OY 121 CGAATTAATTCATCAGAGCTCTATGATCAGTGGAGATGTCCTCCGTGATGTTGATGCC 180
 DB 152 CGAATTAATTCATCAGAGCTCTATGATCAGTGGAGATGTCCTCCGTGATGTTGATGCC 211
 OY 181 AAGGCTTGTGGGCTGTGACTTCTTCTGATGATGGAGATGATCTCAAAATCAAT 240
 DB 212 AAGGCTTGTGGGCTGTGACTTCTTCTGATGATGGAGATGATCTCAAAATCAAT 271
 OY 241 ATCAACCAAGCCCTTGAGCAACAGGTTGAGACGGAAGCTAGAAAAATGTTTCTGTG 300
 DB 272 ATCAACCAAGCCCTTGAGCAACAGGTTGAGACGGAAGCTAGAAAAATGTTTCTGTG 331
 OY 301 ATGACGATGATCTTCAAGAGATCATCCCAAGCCCAACCTCGTGGCAGAGACAT 360
 DB 332 ATGACGATGATCTTCAAGAGATCATCCCAAGCCCAACCTCGTGGCAGAGACAT 391
 OY 361 GTGGTAGTGGCTGTGATAGTATACCAAGAGGTTTCCATTTTCAAGAACCCAGGTT 420
 DB 392 GTGGTAGTGGCTGTGATAGTATACCAAGAGGTTTCTCATTTTCAAGAACCCAGGTT 450
 OY 421 CTCGGCGGTTTTCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGAGGTT 480
 DB 451 CTCGGCGGTTTTCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGAGGTT 510
 OY 481 CGATATGATTTACTGATG 500
 DB 511 CGATATGATTTACTGATG 530

RESULT 11
 AL553746 970 bp mRNA linear EST 16-FEB-2001
 LOCUS AL553746 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1080YF09 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL553746
 VERSION AL553746.1 GI:12893867
 KEYWORDS EST.

SOURCE

ORGANISM

human.

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS

L.M.B., Gruber,C., Jesse,J. and Polayes,D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

CONTACT

Contact: Genoscope

COMMENT

Genoscope Centre National de Sequencage

FEATURES

BP 191 91006 EVRY cedex - France

source

1. 970
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="CS0D1080YF09"
 /clone_lib="LTI_NFL006_PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact: Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax: (1) 301 610 8371
 Email: fliang@life.com URL:
 http://fulllength.invitrogen.com"
 BASE COUNT 222 a 246 c 257 g 240 t 5 others
 ORIGIN

Query Match 97.6%; Score 488; DB 9; Length 970;
 Best Local Similarity 99.8%; Pred. No. 1.4e-140;
 Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 GAATTCCTGAGTCCACAGAGTACAGAAACATTTGCTTTGTCGAGAAAGTCT 60
 DB 225 GAATTCCTGAGTCCACAGAGTACAGAAACATTTGCTTTGTCGAGAAAGTCT 284
 OY 61 CAATCAAAACAACTTACTGAAGTCAAAAGTGTGCGCCCTACATCTCTCAATGTGTT 120
 DB 285 CAATCAAAACAACTTACTGAAGTCAAAAGTGTGCGCCCTACATCTCTCAATGTGTT 344
 OY 121 CGAATTAATTCATCAGAGCTCTATGATCAGTGGAGATGTCCTCCGTGATGTTGATGCC 180
 DB 345 CGAATTAATTCATCAGAGCTCTATGATCAGTGGAGATGTCCTCCGTGATGTTGATGCC 404
 OY 181 AAGGCTTGTGGGCTGTGACTTCTTCTGATGATGGAGATGATCTCAAAATCAAT 240
 DB 405 AAGGCTTGTGGGCTGTGACTTCTTCTGATGATGGAGATGATCTCAAAATCAAT 464
 OY 241 ATCAACCAAGCCCTTGAGCAACAGGTTGAGACGGAAGCTAGAAAAATGTTTCTGTG 300
 DB 465 ATCAACCAAGCCCTTGAGCAACAGGTTGAGACGGAAGCTAGAAAAATGTTTCTGTG 523
 OY 301 ATGACGATGATCTTCAAGAGATCATCCCAAGCCCAACCTCGTGGCAGAGACAT 360
 DB 524 ATGACGATGATCTTCAAGAGATCATCCCAAGCCCAACCTCGTGGCAGAGACAT 583
 OY 361 GTGGTAGTGGCTGTGATAGTATACCAAGAGGTTTCCATTTTCAAGAACCCAGGTT 420
 DB 584 GTGGTAGTGGCTGTGATAGTATACCAAGAGGTTTCTCATTTTCAAGAACCCAGGTT 643
 OY 421 CTCGGCGGTTTTCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGAGGTT 480
 DB 644 CTCGGCGGTTTTCATTTCTCTGAGCGTGTTCAGGGCAGTAGTAGAGAGGTT 703
 OY 481 CGATATGATTTACTGATG 500
 DB 704 CGATATGATTTACTGATG 723

```

RESULT 12
LOCUS      Bf976694
DEFINITION Bf976694 700 bp mRNA linear EST 22-JAN-2001
            602243611f1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:434594 5',
            mRNA sequence.
ACCESSION  Bf976694
VERSION     Bf976694
KEYWORDS    Bf976694.1 GI:12343909
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs.fda.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L10M1204 row: f column: 03
            High quality sequence stop: 665.
FEATURES
    source
        1..700
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:434594"
            /clone_lib="NIH_MGC_48"
            /tissue_type="Primary B-cells from tonsils (cell line)"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: B-cells; Vector: pORF7; Site:1: XhoI;
            Site:2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Size-selected >500bp
            for average insert size 1.8kb. Library constructed by Ling
            Hong in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC library."
BASE COUNT      170 a 173 c 172 g 185 t
ORIGIN
Query Match      97.5%; Score 487.4; DB 10; Length 700;
Best Local Similarity 99.6%; Pred. No. 1.9e-140;
Matches 499; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 GAATTCCTGACTCCACAGAGTGTACAGAAACATTTGCTTTGCTGGAAGCTGCT 60
    |||
DB 138 GAATTCCTGACTCCACAGAGTGTACAGAAACATTTGCTTTGCTGGAAGCTGCT 197
QY 61 CAATATCAAAACATTTACTGAAAGTGAAGTGGTGGCCCTACATCTCAATGGT 120
    |||
DB 198 CAATATCAAAACATTTACTGAAAGTGAAGTGGTGGCCCTACATCTCAATGGT 257
QY 121 CGAATATTAATACATGAGCTCTATGATCAGTGGAGATGTCCTCCGATGATGATGCC 180
    |||
DB 258 CGAATATTAATACATGAGCTCTATGATCAGTGGAGATGTCCTCCGATGATGATGCC 317
QY 181 AAGGCTTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
    |||
DB 318 AAGGCTTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 377
QY 241 ATGACACGAGCCCTGAGGAACAGGTGAGAGGAGGCTGGAAGAAAGTGTGCTG 300
    |||
DB 378 ATGACACGAGCCCTGAGGAACAGGTGAGAGGAGGCTGGAAGAAAGTGTGCTG 437
QY 301 ATGACAGTATCTTCAAGAGTATCCCGACCCACCACTGTTGCCAGAGACAT 360
    |||

```

```

DB 438 ATGACAGTATCTTCAAGAGTATCCCGACCCACCACTGTTGCCAGAGACAT 497
QY 361 GTGATAGTGGCTGTGGATGATGATGATGATGATGATGATGATGATGATGAT 420
    |||
DB 498 GTGATAGTGGCTGTGGATGATGATGATGATGATGATGATGATGATGATGAT 557
QY 421 CTCGGGCTTTTGCATTTCTGAGGCTGTTGAGGAGTGAAGTGAAGTGAAGT 479
    |||
DB 558 CTCGGGCTTTTGCATTTCTGAGGCTGTTGAGGAGTGAAGTGAAGTGAAGT 617
QY 480 TCGATATGATTTACTGCAATTG 500
    |||
DB 618 TCGATATGATTTACTGCAATTG 638

RESULT 13
LOCUS      B685786
DEFINITION B685786 690 bp mRNA linear EST 01-MAY-2001
            602637891f1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4765463 5',
            mRNA sequence.
ACCESSION  B685786
VERSION     B685786
KEYWORDS    B685786.1 GI:13917183
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs.fda.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L10M1624 row: f column: 24
            High quality sequence stop: 686.
FEATURES
    source
        1..690
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4765463"
            /clone_lib="NIH_MGC_48"
            /tissue_type="Primary B-cells from tonsils (cell line)"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: B-cells; Vector: pORF7; Site:1: XhoI;
            Site:2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Size-selected >500bp
            for average insert size 1.8kb. Library constructed by Ling
            Hong in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC library."
BASE COUNT      166 a 172 c 167 g 185 t
ORIGIN
Query Match      97.3%; Score 486.4; DB 10; Length 690;
Best Local Similarity 99.6%; Pred. No. 3.9e-140;
Matches 498; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 GAATTCCTGACTCCACAGAGTGTACAGAAACATTTGCTTTGCTGGAAGCTGCT 60
    |||
DB 137 GAATTCCTGACTCCACAGAGTGTACAGAAACATTTGCTTTGCTGGAAGCTGCT 196
QY 61 CAATATCAAAACATTTACTGAAAGTGAAGTGGTGGCCCTACATCTCAATGGT 120
    |||
DB 197 CAATATCAAAACATTTACTGAAAGTGAAGTGGTGGCCCTACATCTCAATGGT 256
    |||

```

QY 121 CGAATTAATTAATGATGAGTCTATGATGATGAGATGCTCCCTGATGATGATGCC 180
 |||||||
 DB 257 CGAATTAATTAATGATGAGTCTATGATGATGAGATGCTCCCTGATGATGATGCC 316
 |||||||
 QY 181 AAGGCTTTGGGCTGCTGATGCTTCTGCTGATGAGGATGATGATGATGATGATGAT 240
 |||||||
 DB 317 AAGGCTTTGGGCTGCTGATGCTTCTGCTGATGAGGATGATGATGATGATGATGAT 376
 |||||||
 QY 241 ATACACAGAGCCCTTGAGAGACAGAGTGGAGAGAGCTGAGAGAGAGAGAGAGAGAG 300
 |||||||
 DB 377 ATACACAGAGCCCTTGAGAGACAGAGTGGAGAGAGCTGAGAGAGAGAGAGAGAG 436
 |||||||
 QY 301 ATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 |||||||
 DB 437 ATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496
 |||||||
 QY 361 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 |||||||
 DB 497 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
 |||||||
 QY 421 CTCGGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 |||||||
 DB 556 CTCGGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615
 |||||||
 QY 481 CGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
 |||||||
 DB 616 CGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 635
 |||||||

RESULT 14

BI753143

LOCUS 887 bp mRNA linear EST 25-SEP-2001
 DEFINITION 603025996F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:519651 5',
 mRNA sequence.

ACCESSION

BI753143

VERSION

BI753143.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 887)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaab-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L14M11491 row: p column: 24

High quality sequence stop: 809.

Location/Qualifiers

1. 887

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:519651"

/clone_id="NIH_MGC_114"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: EcoRV (destroyed); RNA source anonymous pool of 6

male brains, age range 23-27 yo. Library is 0.190-0.7

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.5 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 019. Note:

this is a NIH_MGC Library."

BASE COUNT 220 a 227 c 209 g 231 t
 ORIGIN

Query Match

Best local similarity 97.3%; Score 486.4; DB 10; Length 887;

Matches 498; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GAATTCGATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 60
 |||||||
 DB 117 GAATTCGATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 176
 |||||||
 QY 61 CAATCAAGAGCAATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 |||||||
 DB 177 CAATCAAGAGCAATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 236
 |||||||
 QY 121 CGAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 |||||||
 DB 237 CGAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 296
 |||||||
 QY 181 AAGGCTTTGGGCTGCTGATGCTTCTGCTGATGAGGATGATGATGATGATGATGAT 240
 |||||||
 DB 297 AAGGCTTTGGGCTGCTGATGCTTCTGCTGATGAGGATGATGATGATGATGATGAT 356
 |||||||
 QY 241 ATACACAGAGCCCTTGAGAGACAGAGTGGAGAGAGCTGAGAGAGAGAGAGAGAGAG 300
 |||||||
 DB 357 ATACACAGAGCCCTTGAGAGACAGAGTGGAGAGAGCTGAGAGAGAGAGAGAGAGAG 416
 |||||||
 QY 301 ATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 |||||||
 DB 417 ATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476
 |||||||
 QY 361 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 |||||||
 DB 477 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 536
 |||||||
 QY 421 CTCGGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 |||||||
 DB 537 CTCGGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 595
 |||||||
 QY 481 CGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
 |||||||
 DB 596 CGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615
 |||||||

RESULT 15

BG333881

LOCUS

602460213F1 NIH_MGC_20

Homo sapiens cDNA clone IMAGE:4576995 5',

mRNA sequence.

ACCESSION

BG333881

VERSION

BG333881.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 936)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaab-r@mail.nih.gov

Tissue Procurement: ATCC/DCTP/DRP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L14M1290 row: j column: 04

High quality sequence stop: 598.

Location/Qualifiers

1. 936

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4576995"

/clone_id="NIH_MGC_20"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: EcoRV (destroyed); RNA source anonymous pool of 6

male brains, age range 23-27 yo. Library is 0.190-0.7

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.5 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 019. Note:

this is a NIH_MGC Library."

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:4576995"  
/clone_1b="NHL MGC-20"  
/tissue_type="melanotic melanoma"  
/lab_host="PH108 (phage-resistant)"  
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT      237 a      235 c      249 g      215 t  
ORIGIN
```

```
Query Match      97.0%: Score 485; DB 10; Length 936;  
Best Local Similarity 99.8%: Pred. No. 1,2e-139;  
Matches 496; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
QY      1 GAATTCCTGACTGCCACAGGTGTACAGAAACATTTGTTGTTGCTGGAAGCTGCT 60  
        |||||||  
Db      51 GAATTCCTGACTGCCACAGGTGTACAGAAACATTTGTTGTTGCTGGAAGCTGCT 110  
        |||||||  
QY      61 CAATCAAAAGACATTTACTGAAGTCAAGTGTGCGCCCTACATCTCTCATGTGGTT 120  
        |||||||  
Db      111 CAATCAAAAGACATTTACTGAAGTCAAGTGTGCGCCCTACATCTCTCATGTGGTT 170  
        |||||||  
QY      121 CGAATATTATACATCAGAGCTCTATCGATCAGTGGAGATGTCTCCGATGATGATGCC 180  
        |||||||  
Db      171 CGAATATTATACATCAGAGCTCTATCGATCAGTGGAGATGTCTCCGATGATGATGCC 230  
        |||||||  
QY      181 AAGGCTTGGTGGGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
        |||||||  
Db      231 AAGGCTTGGTGGGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 290  
        |||||||  
QY      241 ATCACAGAGCCCTTGAGAGAACACAGGTTGAGACGGAAGCTAGAAAAAATGTTCTGTG 300  
        |||||||  
Db      291 ATCACAGAGCCCTTGAGAGAACACAGGTTGAGACGGAAGCTAGAAAAAATGTTCTGTG 350  
        |||||||  
QY      301 ATGACGATGATCTTCAAGAGATCCGCCAGCACCACCAACTGTTGCCACGAGACAAAT 360  
        |||||||  
Db      351 ATGACGATGATCTTCAAGAGATCCGCCAGCACCACCAACTGTTGCCACGAGACAAAT 410  
        |||||||  
QY      361 GTGGTAGTGGTGGATAGTACCAACAACAGGTTCTCCATTTTCAGAAAGCCAGGGGT 420  
        |||||||  
Db      411 GTGGTAGTGGTGGATAGTACCAACAACAGGTTCTCCATTTTCAGAAAGCCAGGGGT 470  
        |||||||  
QY      421 CTCGGGCGTTTTCGATTTCTCTGAGCCGTTTCA-GGGCAGTACGATGAGTGGAGGT 479  
        |||||||  
Db      471 CTCGGGCGTTTTCGATTTCTCTGAGCCGTTTCAAGGGGCACTAGTGAAGTGGAGGT 530  
        |||||||  
QY      480 TCGATATGATTTACTGG 496  
        |||||||  
Db      531 TCGATATGATTTACTGG 547  
        |||||||
```

Search completed: October 10, 2002, 20:27:49
Job time : 702.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 84.9333 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title:	US-09-489-101A-9_COPY_1_500
Perfect score:	500
Sequence:	1 gaattcctgactgcacag.....cgatatgattactgattg 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_032802:★

1	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1980.DAT.*
2	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1981.DAT.*
3	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1982.DAT.*
4	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1983.DAT.*
5	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1985.DAT.*
6	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1985.DAT.*
7	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1986.DAT.*
8	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1987.DAT.*
9	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1988.DAT.*
10	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1990.DAT.*
11	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1990.DAT.*
12	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1991.DAT.*
13	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1992.DAT.*
14	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1993.DAT.*
15	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1994.DAT.*
16	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1995.DAT.*
17	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1995.DAT.*
18	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1997.DAT.*
19	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1998.DAT.*
20	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA1999.DAT.*
21	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA2000.DAT.*
22	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA2001A.DAT.*
23	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA2001B.DAT.*
24	/SIDS1/gcgagata/hold-gemeseq/gemeseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	2291	22	AA01116	Human small cell 1
2	500	100.0	2907	22	AA158623	Human polyucleot
3	498.4	99.7	2968	22	AA160409	Human polyucleot
4	262.2	52.4	2898	22	ABA08061	Human ovarian and
5	262.2	52.4	2898	22	ABA08062	Human ovarian and
6	262.2	52.4	2898	22	AA06787	Human reproductive
7	262.2	52.4	2898	22	AA06788	Human reproductive
8	38.8	7.8	2010	23	ABJ06741	Drosophila melanog
9	38.8	7.8	4272	23	ABJ06740	Drosophila melanog

C	10	38.8	7.8	6162	23	AAI063328	Drosophila melanog
C	11	33.6	6.7	1635	22	AAFE26953	Human cancer assoc
C	12	32.6	6.5	18301	22	AAI098944	Human excretory rel
C	13	32.6	6.5	18301	22	AAI63294	Human kidney relat
C	14	32.4	6.5	936	22	AAFE26310	Pseudomonas sp nod
C	15	32.2	6.4	2187	22	AAAB83052	Human transcritpoid
C	16	32.2	6.4	30078	21	AAAB81520	N. meningitidis pat
C	17	32.2	6.4	349980	21	AAAF21608	Neisseria meningiti
C	18	32	6.4	12156	23	ABLI0434	Drosophila melanog
C	19	31.8	6.4	5151	22	ABAI16681	Human nervous syst
C	20	31.8	6.4	5151	22	AAI05000	Human reproductive
C	21	31.4	6.3	1574	22	AAI04959	Drosophila melanog
C	22	31.4	6.3	3748	23	ABI04958	Drosophila melanog
C	23	31.4	6.3	3813	23	ABLI17884	Drosophila melanog
C	24	30.6	6.1	769	22	AAAI96719	Human neuroblastom
C	25	30.4	6.1	555	22	AAHI0947	Human CDNA clone (
C	26	30.2	6.0	1663	24	AAAS97010	Arabidopsis DNA tor
C	27	30.2	6.0	12138	23	ABLI15540	Drosophila melanog
C	28	30.2	6.0	12161	23	ABLI15542	Drosophila melanog
C	29	30.2	6.0	32476	22	AAAB85314	Human immune/phaem
C	30	30	6.0	710	22	AAAI09376	Human breast cancer
C	31	30	6.0	735	22	AAII17265	Human breast cancer
C	32	30	6.0	1422	22	AAAB89030	Human breast cancer
C	33	30	6.0	3546	20	AAAC90885	CDNA clone PL776-f
C	34	30	6.0	3571	24	AAAO37857	Human peroxisome f
C	35	30	6.0	10899	22	ABAI15344	Human nervous syst
C	36	30	6.0	12758	22	ABAI15345	Human nervous syst
C	37	30	6.0	16066	22	ABAB89022	Escherichia coli p
C	38	29.8	6.0	842	22	AAI193960	Human neuroblastom
C	39	29.8	6.0	1541	16	AAAT04133	Helicobacter pylori
C	40	29.8	6.0	3590	18	AAFE2551	Interleukin-1 rece
C	41	29.8	6.0	4397	23	AAI26422	Drosophila melanog
C	42	29.6	6.0	7957	23	AAI26428	Drosophila melanog
C	43	29.6	5.9	1583	21	AACT78952	Human secreted pte
C	44	29.6	5.9	1677	22	AAI65735	Human retinoblasto
C	45	29.6	5.9	2868	23	AAAS92572	DNA encoding novel

ALIGNMENTS

XX	RESULT 1
XX	AAD1116
ID	AAD1116 strand; DNA; 2291 BP.
AC	AAD1116;
DT	
XX	24-SEP-2001 (first entry)
DE	
XX	Human small cell lung cancer associated gene, eIF2B.
KW	Human: small cell lung cancer; therapy: hCAAP; nucleic acid: NM_001087442.1
KW	melanoma; cancer: colon; breast; head; neck; transitional cancer
KM	leiyosarccoma; eukaryotic translation initiation factor; eIF2B.
KM	sarcoma; cytosolic; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200153349-A2.
XX	
PD	26-JUL-2001.
XX	
PE	19-JAN-2001; 2001WO-USO2015.
FR	
XX	21-JAN-2000; 2000US-0489101.
PA	(LUDW-) LUDWIG INST CANCER RES.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
OR	WPI; 2001-457597/49.

XX Isolated polypeptide, used to treat or prognose a disorder
 PT Characterized by expression of a hCAP e.g. cancer, is encoded by an
 PT Isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 XX Claim 57: Page 95-96; 152pp; English.
 CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is a DNA encoding human eukaryotic translation
 CC initiation factor eIF2B. This small cell lung cancer associated
 CC gene is designated as NY-SCLC-7.
 XX
 SO Sequence 2291 BP; 588 A; 540 C; 630 G; 533 T; 0 other:
 Query Match 100.0%; Score 500; DB 22; Length 2291;
 Best Local Similarity 100.0%; Pred. No. 3.3e-159;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAATTCCTGACTGCCACAGGTGTACAGAAACATTTGCTTTGTTGCTGGAAGCTGCT 60
 DB 1 GAATTCCTGACTGCCACAGGTGTACAGAAACATTTGCTTTGTTGCTGGAAGCTGCT 60
 QY 61 CAATCAAGAAACATTTACTGTAAGTCAAGTGGTGGCCCTGCTGCTGGAAGTGGTT 120
 DB 61 CAATCAAGAAACATTTACTGTAAGTCAAGTGGTGGCCCTGCTGCTGGAAGTGGTT 120
 QY 121 CGAATTAATTAATCAAGCTTATGATGATCTGGAAGATGTCCTCGATGTTGATGCC 180
 DB 121 CGAATTAATTAATCAAGCTTATGATGATCTGGAAGATGTCCTCGATGTTGATGCC 180
 QY 181 AAGGCTTGGTGGCTCTACTTCTTCTGCTGATGAGGATGTCATCTCAATCAATCAAT 240
 DB 181 AAGGCTTGGTGGCTCTACTTCTTCTGCTGATGAGGATGTCATCTCAATCAATCAAT 240
 QY 241 ATACACAGAGCCCTTGAGACAGAGTGTGAGAGGATGAGAAATGTTTCTGTG 300
 DB 241 ATACACAGAGCCCTTGAGACAGAGTGTGAGAGGATGAGAAATGTTTCTGTG 300
 QY 301 ATACAGATGATCTTCAAGAGATCCTCCAGCCCACTGCTGCTGCAAGAGCAAT 360
 DB 301 ATACAGATGATCTTCAAGAGATCCTCCAGCCCACTGCTGCTGCAAGAGCAAT 360
 QY 361 GTGTAGTGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 DB 361 GTGTAGTGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 QY 421 CTCGCGGCTTTTGCAATTTCTCTGAGAGCTGTTTCAAGGAGATGAGATGAGATGAT 480
 DB 421 CTCGCGGCTTTTGCAATTTCTCTGAGAGCTGTTTCAAGGAGATGAGATGAGATGAT 480
 QY 481 CGATATGATTTACTGATG 500
 DB 481 CGATATGATTTACTGATG 500

DE Human polynucleotide SEQ ID NO 826.
 XX
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukaemia; ss.
 XX
 XX Homo sapiens.
 EN WO200153312-A1.
 XX
 XX 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Q, Zhou P, Goodrich R, Drmanac P;
 XX
 XX WPI: 2001-442253/47.
 XX P-PSDB; AAM39467.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 PS Claim 1; SEQ ID NO 826; 10078pp; English.
 XX
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SO Sequence 2907 BP; 734 A; 703 C; 810 G; 653 T; 7 other:
 Query Match 100.0%; Score 500; DB 22; Length 2907;
 Best Local Similarity 100.0%; Pred. No. 3.7e-159;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAATTCCTGACTGCCACAGGTGTACAGAAACATTTGCTTTGTTGCTGGAAGCTGCT 60
 DB 599 GAATTCCTGACTGCCACAGGTGTACAGAAACATTTGCTTTGTTGCTGGAAGCTGCT 658
 QY 61 CAATCAAGAAACATTTACTGTAAGTCAAGTGGTGGCCCTGCTGCTGGAAGTGGTT 120
 DB 659 CAATCAAGAAACATTTACTGTAAGTCAAGTGGTGGCCCTGCTGCTGGAAGTGGTT 718
 QY 121 CGAATTAATTAATCAAGCTTATGATGATCTGGAAGATGTCCTCGATGTTGATGCC 180

Db 719 CGAATTAATTCATGACGCTCATGACACTGAGGAGATGTCCTCCGTGATGTGATGCC 778
QY 181 AAGCTTTGGTGGCTCTGACCTTTCTTGTGTGTATGGGATGTGATCTCAACATCAAT 240
Db 779 AAGCTTTGGTGGCTCTGACCTTTCTTGTGTGTATGGGATGTGATCTCAACATCAAT 838
QY 241 ATCAGCAGACCCCTTGAGAACACAGAGTGTGAGACGAGTGTGAGAAAAAATGTTCTGTG 300
Db 839 ATCAGCAGACCCCTTGAGAACACAGAGTGTGAGACGAGTGTGAGAAAAAATGTTCTGTG 898
QY 301 ATGACGATGATCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 899 ATGACGATGATCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 958
QY 361 GTGAGTGTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 959 GTGAGTGTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1018
QY 421 CTCGCGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 1019 CTCGCGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1078
QY 481 CGATATGATTTACTGATGATG 500
Db 1079 CGATATGATTTACTGATGATG 1098

RESULT 3
AA160409
ID AA160409 standard; cDNA: 2968 BP.
AC AA160409;
XX
XX
XX 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 4398.
XX
XX Human; nontropic; immunosuppressive; cytosolic; gene therapy; cancer;
KW Peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW Amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW Chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukemia; ss.
XX
OS Homo sapiens.
XX
XX MO20015312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000MO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dimauc RT;
XX
XX WPI: 2001-442253/47.
XX P-PSDB: AAM1253.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX

PS Claim 1: SEQ ID NO 4398; 10078bp; English.

XX
XX The invention relates to human nucleic acids (AA15798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM44213) with nontropic,
CC immunosuppressive and cytosolic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilization of the activities such as: immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, hemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX

Sequence 2968 BP; 728 A; 719 C; 825 G; 696 T; 0 other:

Query Match 99.7%; Score 498.4; DB 22; Length 2968;

Best Local Similarity 99.8%; Pred. No. 1,3e-158; Mismatches 1; Indels 0; Gaps 0;

Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATTCCTGACTGCGACAGGTGTACAGGAAACATTTGCTTTGTTGCTGAGAACTGCT 60
Db 262 GAATTCCTGACTGCGACAGGTGTACAGGAAACATTTGCTTTGTTGCTGAGAACTGCT 321
QY 61 CAATCAAGAAAGCAATTTAGTGAAGTCAAAAGTGGTCCGCTACATCTCTCAATGTGTT 120
Db 322 CAATCAAGAAAGCAATTTAGTGAAGTCAAAAGTGGTCCGCTACATCTCTCAATGTGTT 381
QY 121 CGAATTAATTAATGAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Db 382 CGAATTAATTAATGAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 441
QY 181 AAGCTTTGGTGGCTCTGACCTTTCTTGTGTGTATGGGATGTGATCTCAACATCAAT 240
Db 442 AAGCTTTGGTGGCTCTGACCTTTCTTGTGTGTATGGGATGTGATCTCAACATCAAT 501
QY 241 ATCAGCAGACCCCTTGAGAACACAGAGTGTGAGACGAGTGTGAGAAAAAATGTTCTGTG 300
Db 502 ATCAGCAGACCCCTTGAGAACACAGAGTGTGAGACGAGTGTGAGAAAAAATGTTCTGTG 561
QY 301 ATGACGATGATCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 562 ATGACGATGATCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
QY 361 GTGAGTGTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 622 GTGAGTGTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 681
QY 421 CTCGCGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 682 CTCGCGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 741
QY 481 CGATATGATTTACTGATGATG 500
Db 742 CGATATGATTTACTGATGATG 761

RESULT 4
ABA08061/C
ID ABA08061 standard; DNA: 2898 BP.
AC ABA08061;
XX
XX
XX 11-JAN-2002 (first entry)
XX
XX Human ovarian and breast cancer associated polynucleotide SEQ ID NO 856.
XX
XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
KW

08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488786/53.
XX
XX New isolated ovarian and/or breast cancer related nucleic acids and
PT polypeptides, useful for diagnosing, treating and/or preventing human
PT diseases and disorders, particularly ovarian and/or breast cancer -
XX
PS Disclosure: SEQ ID NO 856; 577bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA07454-ABA08224) and proteins
CC (ABA10743-ABA10980) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful.
CC In the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2898 BP; 796 A; 744 C; 600 G; 758 T; 0 other:
Query Match 52.4%; Score 262.2; DB 22; Length 2898;
Best Local Similarity 79.4%; Pred. No. 2.8e-78;
Matches 366; Conservative 0; Mismatches 3; Indels 92; Gaps 1;

RESULT 5
ABA08062/c
ID ABA08062 standard; DNA; 2898 BP.
XX
XX ABA08062;
AC
XX
XX 11-JAN-2002 (first entry)
DE
XX
XX Human ovarian and breast cancer associated polynucleotide SEQ ID NO 857.
XX
XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; ds.
XX
OS Homo sapiens.
XX
XX WO200155325-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01345.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
XX
XX 14-AUG-2000; 2000US-0224519.
XX
XX 14-AUG-2000; 2000US-0225213.
XX
XX 14-AUG-2000; 2000US-0225214.
XX
XX 14-AUG-2000; 2000US-0225266.
XX
XX 14-AUG-2000; 2000US-0225267.
XX
XX 14-AUG-2000; 2000US-0225268.
XX
XX 14-AUG-2000; 2000US-0225447.
XX
XX 14-AUG-2000; 2000US-0225470.
XX
XX 14-AUG-2000; 2000US-0225757.
XX
XX 14-AUG-2000; 2000US-0225758.
XX
XX 14-AUG-2000; 2000US-0225759.
XX
XX 14-AUG-2000; 2000US-0225759.
XX
XX 18-AUG-2000; 2000US-0226279.
XX
XX 22-AUG-2000; 2000US-0226681.
XX
XX 22-AUG-2000; 2000US-0226682.
XX
XX 23-AUG-2000; 2000US-0227009.
XX
XX 30-AUG-2000; 2000US-0228924.
XX
XX 01-SEP-2000; 2000US-0229287.
XX
XX 01-SEP-2000; 2000US-0229343.
XX
XX 01-SEP-2000; 2000US-0229344.
XX
XX 01-SEP-2000; 2000US-0229345.
XX
XX 05-SEP-2000; 2000US-0229509.
XX
XX 05-SEP-2000; 2000US-0229513.
XX
XX 06-SEP-2000; 2000US-0230437.
XX
XX 06-SEP-2000; 2000US-0230438.
XX
XX 08-SEP-2000; 2000US-0231242.
XX
XX 08-SEP-2000; 2000US-0231243.
XX
XX 08-SEP-2000; 2000US-0231244.
XX
XX 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232387.
 PR 14-SEP-2000; 2000US-0232598.
 PR 14-SEP-2000; 2000US-0232599.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234597.
 PR 25-SEP-2000; 2000US-0234598.
 PR 26-SEP-2000; 2000US-0234584.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0246177.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-488786/53.
 XX
 PT New isolated ovarian and/or breast cancer related nucleic acids and
 PT polypeptides, useful for diagnosing, treating and/or preventing human
 PT diseases and disorders, particularly ovarian and/or breast cancer -
 XX
 PS Disclosure; SEQ ID NO 857; 577bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA07454-ABA08224) and proteins
 CC (ABA07454-ABA0980) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 2898 BP; 796 A; 744 C; 600 G; 758 T; 0 other;
 Query Match 52.4%; Score 262.2; DB 22; Length 2898;
 Best Local Similarity 79.4%; Pred. No 2.8e-78;
 Matches 366; Conservative 0; Mismatches 3; Indels 92; Gaps 1;
 QY 76 TTTACTGATCAAGTGGTGGCCGCTCATCTGATGTTGGTCAATATACATCA 135
 Db 644 TTTTGAAGTCAAGTGGTGGCCGCTCATCTGATGTTGGTCAATATACATCA 585
 QY 136 GAGCTATGATCACTGGAGATGCTCGTGATGTTGATGTCAGAGCTTGGTCCG 195
 Db 584 GAGCTATGATCACTGGAGATGCTCGTGATGTTGATGTCAGAGCTTGGTCCG 525
 QY 196 TCTGACTTCTTCTGTTGATGGGATGTCATCAATCAATCATATACAGAGCCCTT 255
 Db 524 TCTGACTTCTTCTTGGTGTATGGGATGTCATCAATCAATCAATACAGAGCCCTT 465
 QY 256 GAGGACAC----- 264
 Db 464 GAGGACACAGTCAAGATGGAATAAGACAGGTTAAAGACACGAGAGCC 405
 QY 265 -----AGTTGAGACGAGACTAG 283
 Db 404 TGACACTGTTTTTTCAGTCTGTCCTCTGTCGCTTTATAGTGTAGAGGAGACTAG 345

Db 224 TTCAGAGACCCAGGCTCTCCGGCTTTGCATTCTCTG 184

RESULT 8

ABL06741

ID ABL06741 standard; cDNA; 2010 BP.

XX ABL06741;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 14705.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX P-PSDB; ABB62638.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 14705; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2010 BP; 463 A; 548 C; 596 G; 403 T; 0 other;

Query Match

Best Local Similarity 7.8%; Score 38.8; DB 23; Length 2010;

Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

XX

XX

XX

XX

XX

XX

ID ABL06740 standard; cDNA; 4272 BP.

XX ABL06740;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 14702.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX P-PSDB; ABB62637.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 14702; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4272 BP; 1069 A; 1157 C; 1092 G; 954 T; 0 other;

Query Match

Best Local Similarity 7.8%; Score 38.8; DB 23; Length 4272;

Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 9

ABL06740/c

ID ABL06740 standard; cDNA; 6162 BP.

XX ABL06740;

XX 26-MAR-2002 (first entry)

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13466.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX MO20017042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001MO-US09231.
PE
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW.
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX P-PSDB; ABB62225.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PI interactions -
XX
XX Claim 1; SEQ ID NO 13466; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
CC
XX
XX Sequence 6162 BP; 1591 A; 1658 C; 1564 G; 1349 T; 0 other;
SQ
Query Match 7.8%; Score 38.8; DB 23; Length 6162;
Best Local Similarity 51.8%; Pred. No. 0.041;
Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
OY 118 GTTCGATATTTACATCAGACTCTATGATCGACTGGAGATGTCCTCGATGTTGAT 177
DB 422 GTCCATGCTATGCGCGAGGGTGTGCTGCTGCGCGCATGCCATGCTGATTTGGAC 363
OY 178 GCCAAGCCTTTGGCGCTCTACTTCTCTGCTGATGATGATGATCATAACATC 237
DB 362 AATAGAGCGCTGATGCGCGATCTGCTGCTGCGCGCATGAGGAGCAATGCGC 303
OY 238 AATATCACAGAGCCTTGAAGACAGATTTGAGACAGAGAGAGAGAA 287
DB 302 GATCTGCGCCCACTGCTGAGACAGCAAGCGCGCAAGTTCGACAA 253
RESULT 11
AA26953/c
ID AA26953 standard; cDNA; 3635 BP.
XX
XX AA26953;
AC
XX
XX 10-APR-2001 (first entry)
DT
XX
XX Human cancer associated antigen precursor HOM-TES-84/6 cDNA SEQ ID NO:5.
DE
XX
XX Human; cancer associated antigen precursor; cancer associated antigen;
KW seminoma; HLA; human leukocyte antigen; cytosolic; gene therapy;
KM

KM vaccine; cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200100874-A2.
PN
XX
XX 04-JAN-2001.
PD
XX
XX 23-JUN-2000; 2000MO-US17207.
PE
XX
XX 30-JUN-1999; 99US-0346498.
PR
XX
XX (LUDWIG) LUDWIG INST CANCER RES.
PA
XX
XX Sahn U, Tureci O, Pfreundschuh M.
PI
XX
XX WPI; 2001-112465/12.
DR
XX
XX P-PSDB; AAB75607.
XX
XX Diagnosing a disorder characterized by expression of a human cancer
PT associated antigen precursor, comprises detecting interaction of an
PI agent with a nucleic acid molecule encoding the antigen precursor -
XX
XX Claim 48; Page 90-91; 126pp; English.
PS
XX
XX The present invention describes a method for diagnosing a disorder
CC characterised by expression of a human cancer associated antigen (CMA)
CC precursor (I) coded by a NA Group 1 nucleic acid molecule (NI)
CC comprising contacting the biological sample with an agent (A) that
CC specifically binds to NI, (I) or its fragment, complexed with an human
CC leukocyte antigen (HLA) molecule and determining the interaction between
CC the agent and NI or (I). (I) has cytotoxic activity and can be used in
CC gene therapy and vaccine production. The method can be used for treating
CC a subject with a condition characterised by expression of (I) in cells
CC of a subject. The present sequence represents a human cancer associated
CC antigen precursor cDNA sequence which is used in the exemplification of
CC the present invention.
XX
XX
XX Sequence 3635 BP; 1162 A; 741 C; 797 G; 935 T; 0 other;
SQ
Query Match 6.7%; Score 33.6; DB 22; Length 3635;
Best Local Similarity 55.0%; Pred. No. 1.8;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
OY 21 TGNACGGAACATTTGCTTTGCTGGAAGCTGCTCAATCAAGACATTACT 80
DB 559 TGCTTAGGACCACTTCTATGTTGTGCTCAGCGAATAATCTTAAATTTCGA 500
OY 81 GAAGTCAAGGTGTCGCCCTACATCTCTCAATGTCGTTCAATTAATTCAGACT 140
DB 499 GATTTGAAGTAGCACCACTTAATTTCTTTGCTTGTGCACTGTTCAATGAGACT 440
RESULT 12
AA198944/c
ID AA198944 standard; DNA; 18301 BP.
XX
XX AA198944;
AC
XX
XX 07-JAN-2002 (first entry)
DT
XX
XX Human excretory related polynucleotide SEQ ID NO 708.
DE
XX
XX Human; excretory; neuroprotective; cytosolic; dermatological; vitruclide;
KW immunosuppressive; anti-inflammation; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antischizoid; antianemic; antiarthritic; cancer;
KW antipneumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW excretory system; ds.
XX
XX
XX Homo sapiens.
OS

PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465569/50.
DR

PT Isolated nucleic acid molecule encoding excretory system antigen is
PT used in preventing, treating or ameliorating a medical condition -
XX
XX
PS Example 2; SEQ ID NO 708; 574bp + Sequence Listing; English.

CC The invention relates to novel excretory system related human
CC polynucleotides (AA198567-AA199503) and the encoded proteins
CC (AA199594-AA199613) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy, especially
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, diabetes mellitus, Crohn's
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral ischaemia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 18301 BP; 5076 A; 4392 C; 3842 G; 4991 T; 0 other;

Query Match 6.54; Score 32.6; DB 22; Length 18301;
Best Local Similarity 54.64; Pred. No. 9.5; Indels 0; Gaps 0;
Matches 65; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 133 TCAGAGCTTATCGATCAGTGGAGATGCTCCGCGTGTGATGTCAGAGCTTTGTG 192
DB 198 TCCGGGCTTCAATAGTCTTGTATGATGTTGAATTTGATGATGATTTGTG 139
OY 193 CGCTCGATTTCTTCTGTATGAGGATGTCATCAACATCAATATTCACAGAGC 251
DB 138 AACTGTGCTTTAGCTTTCATTAATAAGCTTTCTTCAATCAATATCTCTGTC 80

RESULT 13
AA163294/C
ID AA163294 standard; DNA; 18301 BP.
XX
AC AA163294;
XX

XX 22-OCT-2001 (first entry)
DE Human kidney related polynucleotide SEQ ID NO 609.
XX
XX

KW Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
KW optalmological; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
XX

OS Homo sapiens.
XX
XX

PN WO200153323-A2.
XX
XX

PD 02-AUG-2001.
XX
XX

PF 17-JAN-2001; 2001WO-US01343.
XX
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0186874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0196123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218299.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PF New DNA encoding a bacterial nodulation protein, useful e.g. for
PT promoting root formation and yield of plants, and producing transgenic
XX plants -
XX
XX
PS Claim 3a: Page 14; 16pp; German.
XX
CC This invention describes novel DNA sequences (I) that express products
CC having the biological function of nodulation factors. The invention also
CC describes (1) recombinant expression vectors containing (1); (2) or
CC prokaryotic and eukaryotic cells transformed or transfected with (1) or
CC the vector of (1); (3) production of nodulation products (II) of (1), and
CC cells of (2); (4) (partial) expression products (II) of (1), and
CC synthetic proteins or peptides with the same sequences; (5) monoclonal or
CC polyclonal antibodies (Ab) specific for (II); (6) hybridoma cells that
CC produce monoclonal Ab; and (7) transgenic plants that contain cells of
CC (2), (1), and their fragments, are useful for expression of recombinant
CC nodulation factors, are useful as probes and primers for detection,
CC isolation and amplification of full-length cDNA sequences and are used to
CC produce transgenic plants. Cells transformed with (1), and recombinant
CC nodulation factors or synthetic peptides or proteins with the same
CC activities, are used to promote growth, development and yield of plants,
CC particularly leguminosae. The nodulation factors encoded by (1) promote
CC root formation in plants and improve yield, particularly when grown on
CC soil contaminated with xenobiotics injurious to plants.
XX
SQ Sequence 936 BP; 182 A; 293 C; 279 G; 182 T; 0 other:
XX
Query Match 6.5%; Score 32.4; DB 22; Length 936;
Best Local Similarity 52.2%; Pred. No. 2.3;
Matches 72; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 343 CGTTCGACGAGACATGCTGTGCTGTGATACACAAACAGGCTCTCCAT 402
DB 312 CGAGCCACACAGGAGGTGCTGCTGCGCGCCGATGCAACACGCGCTGGCCCT 253
QY 403 TTTTCAGAGACCCAGGCTCTGCGCTTTTGCATTTCTCTGAGCGCTTTTCAGGCGCAGT 462
DB 253 TATCGAGCGCGCAATGTGACAGACACCTCGCGTACCGCTTGCCAGTCCAGACATA 193
QY 463 AGTGATGGAGTGGAGGCT 480
DB 192 GGGAGTGGGCTGCATCTT 175
XX
RESULT 15
ABA83052/C
ID ABA83052 standard; DNA: 2187 BP.
XX
AC ABA83052;
XX
DT 05-FEB-2002 (first entry)
XX
DE Human transcription factor TRFX-79 coding sequence.
XX
XX Human: transcription factor; TRFX; cell proliferative disease;
KM autoimmune disease; inflammation; neurological disease;
KM developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
XX neuroprotective; anti-inflammatory; gene therapy; ds.
XX
OS Homo sapiens.
XX
XX WO200172777-A2.
PN
XX
PD 04-OCT-2001.
XX
PF 13-MAR-2001; 2001WO-US08117.
XX
PR 13-MAR-2000; 2000US-0188986.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JU, Baughn MR, Yue H, Lal P, Lu DM, Patterson C;
PI Azimzai Y, Bandman O, Tang YT, Matchur P, Shah P, Au-Young J;

PI Reddy R;
XX
XX WPI, 2001-570896/64.
DR
DR P-PDB; ABB50228.
XX
XX
PT Novel transcription factor polypeptides, used to treat diseases
PT associated with altered activity and expression of TRFX, and to screen
PT for agents capable of modulating its activity -
XX
PS Claim 11: Pages 307-308; 327pp; English.
XX
XX The present sequence is the coding sequence for a human transcription
XX factor. The transcription factor and its coding sequence are useful in
XX the diagnosis, treatment and prevention of diseases associated with
XX altered expression of the transcription factor e.g. cell proliferative,
XX autoimmune/inflammatory, neurological and developmental disorders. A
XX number of specific disorders/diseases are given in the specification,
XX including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
XX allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
XX dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
XX Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
XX psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
XX colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
XX disease, stroke, and viral, bacterial, fungal and protozoal infections.
XX
SQ Sequence 2187 BP; 386 A; 697 C; 737 G; 367 T; 0 other:
XX
Query Match 6.4%; Score 32.2; DB 22; Length 2187;
Best Local Similarity 57.4%; Pred. No. 4.2;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 378 TAGTACACAAACAGGCTTCTCCATTTTCAGAGACCCAGGCTCCGGCTTTGCATT 437
DB 788 TGGCAGCACTCCAGGGGTCTGACGCTTGAGAACCCGCCCTGCCAGCGCTGGCATG 729
QY 438 TCTCTGAGGCTGTTTCAGGCGAGTAGTAGTAGGAGG 478
DB 728 TCGGTTGCACTGCTCGATGCAAGGTGAGAGAGAGAG 688
XX
Search completed: October 10, 2002, 15:15:35
Job time : 118 secs

Best Local Similarity 5.4%; Pred. No. 1.1e-06;
Matches 17; Conservative 180; Mismatches 120; Indels 0; Gaps 0;
QY 135 GTGGAGAGATGTTGAAGCTTTTGAAGAGAGAAACAAATGCTTGTCTTTT 194
1463 GTAGTAAAGAGATAGAAATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRR 1404
QY 195 CAGGAACCTGAGATAGAAAGATTTCGAGATCAGAGATGAAGTGAAGTCACTG 254
1403 RRR 1344
QY 255 GAAGAGTTCCATGCTGAAACATCATGATCCAGAGTGAAGCAAGATCATTA 314
1343 RRR 1284
QY 315 GCAGTTCTGACTGCGCTGTGATACAGGCTACAGAGAGATCAAGCAAGCT 374
1283 RRR 1224
QY 375 CATAAAGCAATGTTTAAACATCACCAGACAAAGCAAGAGCTGTGAACCAATA 434
1223 RRR 1164
QY 435 AAAGAGAGATATGA 451
DB 1163 RRR

RESULT 2
US-08-346-849-1
; Sequence 1, Application US/08346849
; Patent No. 5670483
; GENERAL INFORMATION:
; APPLICANT: Zhang, Shuguang
; APPLICANT: Lockshin, Curtis
; APPLICANT: Rich, Alexander
; APPLICANT: Holmes, Todd
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,849
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28 DECEMBER 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOK, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1292..2590
; OTHER INFORMATION: /product= "zuotin"
US-08-346-849-1
Query Match 10.4%; Score 51.8; DB 1; Length 3083;
Best Local Similarity 55.7%; Pred. No. 1.6e-06;
Matches 123; Conservative 0; Mismatches 92; Indels 6; Gaps 1;
QY 276 ACACTTGATCCCAAGACTGGAAGAACAGATCATTAATGCACTTGTGACCTGGCCAT 335
1550 ACTGATGATGCTGAGATGGAAGAACTGCGGATTTGATGCTGATGCTGATGCTTAAG 1609
QY 336 GTGAGATAGAGGCTGACAGAGAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 395
1610 TTGCGTTTCAGAGCTACTGAAATCATATCATCAAGCTTCACAGAAACAGTTGTCAAG 1669
QY 396 CATACCCAGACCAAG 455
1670 TACCATCCAGACCAAG 1723
QY 456 TTGACTTGCACTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 496
DB 1724 TTCAAGATTATTCAAAAGGCTTTGAAACTTTGACTGATTC 1764

RESULT 3
US-08-293-284A-1
; Sequence 1, Application US/08293284A
; Patent No. 5955343
; GENERAL INFORMATION:
; APPLICANT: Holmes, Todd
; APPLICANT: Zhang, Shuguang
; APPLICANT: Rich, Alexander
; APPLICANT: DiPersio, C. Michael
; APPLICANT: Lockshin, Curtis
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/293,284A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOK, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double


```

CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/309,682
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/007,484
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10081
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1293 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-309-682-1

Query Match          7.0%; Score 34.8; DB 4; Length 1293;
Best local similarity 53.7%; Pred. No. 0.25;
Matches    72; Conservative      0; Mismatches   62; Indels     0; Gaps       0;

QY  194 TCAGGAACGCGAGATTAAGAACAATTCTCCGAGATTCGAAAGATCAAGATAAATTGCAGTT 253
Db   213 TCAGGATTCAGAACGCCACTGTAGGGCTTCTCAAGAAAGAAATTAAGCATTTACCAAATTC 272
QY   254 GGAAGATTTCCCATGCTGAAAACACTTGATCCCAGAACTGGAAGAACCAACAGATCATTA 313
Db   273 AGAAGATTAACCGAGAGAGAAAGACGTTGACACGAGAAAGAACTATAGAGAAAAATTATTC 332
QY   314 TGCA GTTCTTGAC 327
Db   333 TGAAGTTCTTGAAC 346

RESULT 8
US-08-961-083-159
Sequence 159, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
```



```

TELEX: 248633
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Borrelia garinii
STRAIN: IP90
IMMEDIATE SOURCE:
CLONE: pJB-101
FEATURE:
NAME/KEY: CDS
LOCATION: 380..2245
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 380..442
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 443..2242
US-08-262-220-7

```

Query Match Best Local Similarity 56.8%; Score 34.4; DB 3; Length 2547;

Matches 84; Conservative 0; Mismatches 61; Indels 3; Gaps 1;

```

QY 342 TACAAAGCTACACAGACAGATCAAGAGCTCATTAAGCAATGTTTAAACATCAC 401
    || || || || || || || || || || || || || || || || || || || ||
DB 2472 TATAGCTCTGATGATGATGATTAAGAGCTTATTAAGCTTATTAAGATTCAT 2413
QY 402 CCAGCAAGAGGAAGAGCTGTGATCAATTAAGAGAGATATGACTTCACT 461
    || || || || || || || || || || || || || || || || || || || ||
DB 2412 CCTGACA--GGTTGGCAATGATCCTGTAGAGCAAAAAGATGCAATGATTAATT 2356
QY 462 TGCATTAAGCTTATGAAATGTTAT 489
    || || || || || || || || || || || || || || || || || || || ||
DB 2355 AAATTCAGATGCTTATGAAAAGATT 2328

```

RESULT 11

US-08-471-733-7/c

; Sequence 7, Application US/08471733

; Patent No. 6068842

; GENERAL INFORMATION:

; APPLICANT: BERGSTROM SVEN

; APPLICANT: BARBOUR ALAN G.

; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELLIA

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 SEVENTH STREET, N.W.

; CITY: WASHINGTON

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/08/471.733

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: 08/262,220

; FILING DATE: 20-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: COOPER, IVER P.

; REGISTRATION NUMBER: 28,005

; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Borrelia garinii
STRAIN: IP90
IMMEDIATE SOURCE:
CLONE: pJB-101
FEATURE:
NAME/KEY: CDS
LOCATION: 380..2245
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 380..442
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 443..2242
US-08-471-733-7

```

Query Match Best Local Similarity 56.8%; Score 34.4; DB 3; Length 2547;

Matches 84; Conservative 0; Mismatches 61; Indels 3; Gaps 1;

```

QY 342 TACAAAGCTACACAGACAGATCAAGAGCTCATTAAGCAATGTTTAAACATCAC 401
    || || || || || || || || || || || || || || || || || || || ||
DB 2472 TATAGCTCTGATGATGATGATTAAGAGCTTATTAAGCTTATTAAGATTCAT 2413
QY 402 CCAGCAAGAGGAAGAGCTGTGATCAATTAAGAGAGATATGACTTCACT 461
    || || || || || || || || || || || || || || || || || || || ||
DB 2412 CCTGACA--GGTTGGCAATGATCCTGTAGAGCAAAAAGATGCAATGATTAATT 2356
QY 462 TGCATTAAGCTTATGAAATGTTAT 489
    || || || || || || || || || || || || || || || || || || || ||
DB 2355 AAATTCAGATGCTTATGAAAAGATT 2328

```

RESULT 12

US-08-468-878-7/c

; Sequence 7, Application US/08468878

; Patent No. 6090586

; GENERAL INFORMATION:

; APPLICANT: BERGSTROM SVEN

; APPLICANT: BARBOUR ALAN G.

; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELLIA

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 SEVENTH STREET, N.W.

; CITY: WASHINGTON

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/08/468,878

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: 08/262,220

; FILING DATE: 20-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: COOPER, IVER P.

; REGISTRATION NUMBER: 28,005

```

REFERENCE/DOCKET NUMBER: BERGSTROM=3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Borrelia garinii
STRAIN: Ip90
IMMEDIATE SOURCE:
CLONE: pJB-101
FEATURE:
NAME/KEY: CDS
LOCATION: 380..2245
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 380..442
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 443..2242
US-08-468-878-7

Query Match      6.9%; Score 34.4; DB 3; Length 2547;
Best Local Similarity 56.8%; Pred. No. 0.49;
Matches 84; Conservative 0; Mismatches 61; Indels 3; Gaps 1

QY   342 TACAAGCTACACAGACGATCAAAAGCAGCTCATATAAGCAATGTTTTAAAACATCAC 401
    ||| |||| | |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   2472 TATAGTGCTAGTAGATGATGATTAATAAAGGCTATTATAAAGTCGTGTATATAAATATCAT 2413

QY   402 CCAGACAAAGCGAAAGCAGCTGCTGAACCAATAAAGAAGAGATATGACTACTGCAC 461
Db   2412 CCTGACA--GGTTGCCAATGATCTCTTAAGGCCAAAAAGATGCAATGATTAATTTATT 2356

QY   462 TGCATACTAAGCTTATGAATGTTAT 489
    ||| |||| |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   2355 AAAATTCAGATGCTTATGAAGAATTT 2328


RESULT 13
US-08-750-494-7/c
Sequence 7, Application US/08750494
Patent No. 6204018
GENERAL INFORMATION:
APPLICANT: BERGSTROM SVEN
APPLICANT: BARBOUR ALAN G.
TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELLIA
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 SEVENTH STREET, N.W.
CITY: WASHINGTON
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0., Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,494
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/262,220
FILING DATE:
ATTORNEY/AGENT INFORMATION:
```

```

NAME: COOPER, IVER P.
REGISTRATION NUMBER: 28, 005
REFERENCE/DOCKET NUMBER: BERGSTROM=3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
Best local similarity 56.8%; Pred. No. 0.49;
ORIGINAL SOURCE:
ORGANISM: Borrelia garinii
STRAIN: Ip90
IMMEDIATE SOURCE:
CLONE: pJB-101
FEATURE:
NAME/KEY: CDS
LOCATION: 380..2245
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 380..442
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 443..2242
US-08-750-494-7

Query Match 6.9% Score 34.4; DB 4; Length 2547;
Best local similarity 56.8%; Pred. No. 0.49;
Matches 84; Conservative 0; Mismatches 61; Indels 3; Gaps 1.

QY 342 TACAGGCTCACAGAGACAGATCAAAAGCAGCTCATTAAGCAATGTTTAAAACATCAC 401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2472 TATNGTCCTAGTATGTATGATGAGATTAAAAAGCTTTATAAAGCTTGTTATTAATATCAT 2413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 402 CCAGACAACGGAACAGCAGCTGGTAGAACCAATAAAAAGAAGAGATATGACTTCTACT 461
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2412 CCTGACA---GGTTTCCAATGATCCCTGTAAGGCCAAAAAGATGCAAATGATTAATTATT 2356
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 462 TGCATACCTAAAGCTTATGAATGTTAT 489
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2355 AAAATTCAAGATGCTTATGAAAAGATT 2328
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-106-638-1
Sequence 1, Application US/09106638
Patent No. 6093556
GENERAL INFORMATION:
APPLICANT: kanji NAKAMURA
APPLICANT: hitoaki ISHIDA
TITLE OF INVENTION: GENE RECOMBINANT WITH BIODEGRADABILITY
TITLE OF INVENTION: FOR CHLORINATED ETHYLENE AND BIO-TREATMENT OF
TITLE OF INVENTION: CHLORINATED ETHYLENE THEREWITH
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: c/o KURITA WATER INDUSTRIES LTD.
STREET: 4-7, Nishi-Shinjuku 3-Chome
City: Shinjuku-Ku
STATE: Tokyo
COUNTRY: Japan
ZIP: 160-0023
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.5 inches, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,638
FILING DATE:

```

```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 174996/1997, Japan
FILING DATE: 30-June-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 174997/1997, Japan
FILING DATE: 30-June-1997
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4800 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORGANISM: Pseudomonas putida
INDIVIDUAL ISOLATE: KMI-9
FEATURE:
NAME/KEY: peptide
LOCATION: 127..345
IDENTIFICATION METHOD: E
OTHER INFORMATION: phe2 of phenol-hydroxylase
FEATURE:
NAME/KEY: peptide
LOCATION: 434..1429
IDENTIFICATION METHOD: E
OTHER INFORMATION: pheA of phenol-hydroxylase
FEATURE:
NAME/KEY: peptide
LOCATION: 1440..1712
IDENTIFICATION METHOD: E
OTHER INFORMATION: pheB of phenol-hydroxylase
FEATURE:
NAME/KEY: peptide
LOCATION: 1754..3268
IDENTIFICATION METHOD: E
OTHER INFORMATION: pheC of phenol-hydroxylase
FEATURE:
NAME/KEY: peptide
LOCATION: 3301..3660
IDENTIFICATION METHOD: E
OTHER INFORMATION: pheD of phenol-hydroxylase
FEATURE:
NAME/KEY: peptide
LOCATION: 3689..4756
IDENTIFICATION METHOD: E
OTHER INFORMATION: pheE of phenol-hydroxylase
US-09-106-638-1

Query Match
Best Local Similarity 6.6%; Score 33.2; DB 3; Length 4800;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 36 GGGCGCATCATGCTGCTTGCACAGCGCGGCGGACACCGGCATCACCAC 95
DB 827 GCGGCCAACATGACACATCTCTATGCGCGGACGCTGCGGCGGACGCGGCATCACCAC 886
QY 96 GCTGTGACCT 105
DB 887 GCGTGCACCT 896

```

```

RESULT 15
US-08-681-129-1
Sequence 1, Application US/08681129
Patent No. 573854
GENERAL INFORMATION:
APPLICANT: Mettenleiter, Thomas Cristoph
TITLE OF INVENTION: Pseudorabies virus vaccine
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: Organon Teknika Corporation
STREET: 1330 Piccard Drive

```

```

CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,129
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/244,446
FILING DATE: 02-JUN-1994
CLASSIFICATION: 424
APPLICATION NUMBER: EP 92.203.079.6
FILING DATE: 06-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1578 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
STRAIN: Kaplan
US-08-681-129-1

Query Match
Best Local Similarity 6.6%; Score 33; DB 1; Length 1578;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 8 GAGCGCTGGCGCCACCGGCTAGACCGGCGGCATCATCTCTCTGCGACGCGCGC 67
DB 774 GATGCTGTGCGACCTGTGCGCGGCGGACGCGCATCTCTGCGGCGCGCGCTC 833
QY 68 GAGCGCGCGGCGCGCGCATCACCACGCTGTGACCTTGCCCTC 112
DB 834 GAGCAGCGCTGCGCGCGCTGTGCGAGGCGCGTCCGCGCGTGC 878

```

Search completed: October 10, 2002, 18:14:08
Job time : 31.7333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds
(without alignments)
9699.805 Million cell updates/sec

Title: US-09-489-101a-8_COPY_1_500

Perfect score: 500
Sequence: 1 gggagcgtgagcgcgtgcgc.....aaatgtatctgacacgtg 500

Scoring table: IDENTITY_MQC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estchum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	761	10	BF037074 601456938
2	500	100.0	844	9	AU121047
3	500	100.0	920	11	BC012376 Homo sapi
4	489	97.8	666	10	BC0120251
5	489	97.8	780	9	AL548837
6	488	97.6	689	10	BI460193
7	488	97.6	937	10	BC396661
8	486	97.3	955	9	AL517760
9	485.4	97.1	829	10	BC718443
10	483.2	96.6	665	9	AV655598
11	476	95.2	859	10	BC165721
12	472.8	94.6	735	9	AI364977
13	460.6	92.1	721	9	AW161245
14	438.2	87.6	656	10	AU128589
15	422.4	84.5	711	10	BF562033
16	413	82.6	700	10	BF061885
17					BE865243

18	413	82.6	721	10	BE865164	BE865164	UI-M-BH1-
19	403.6	80.7	1014	11	AK013295	AK013295	Mus muscu
20	400.2	80.0	718	10	BG503926	BG503926	602552823
21	384.8	77.0	759	9	AM107566	AM107566	u191e01.y
22	369	73.8	460	10	BF602672	BF602672	268056.MA
23	354.6	70.9	723	10	BG721900	BG721900	602695624
24	349.4	69.9	774	9	AU117184	AU117184	602794978
25	344.4	68.2	1073	10	BG861987	BG861987	602794870
26	341.2	68.0	815	10	BG548703	BG548703	602576314
27	340	68.0	689	10	BG548703	BG548703	602576314
28	332.8	66.6	525	10	BF719381	BF719381	mab39g04.
29	332.4	66.5	449	9	AA672310	AA672310	v889g09.s
30	328.2	65.6	459	10	BF593004	BF593004	naa30b10.
31	327.4	65.5	477	9	A1876355	A1876355	uk74a02.y
32	325.8	65.2	489	9	A1931324	A1931324	u167q11.y
33	325.8	65.2	624	10	BE956852	BE956852	UI-M-BG2-
34	322.2	64.4	597	9	AM270086	AM270086	xv37q10.x
35	322.2	64.4	613	10	BG145510	BG145510	uu89a02.x
36	294.6	58.9	508	10	BM488205	BM488205	pgm28.pKO
37	274	54.8	604	10	BE679321	BE679321	dc6109.y
38	254.2	50.8	414	9	A1175585	A1175585	EST219140
39	252	50.4	574	9	AM612011	AM612011	hg83e12.x
40	241.8	48.4	804	10	BF167144	BF167144	601774707
41	240.2	48.0	743	9	AM174404	AM174404	f142g10.y
42	239.8	48.0	610	10	BJ003633	BJ003633	602913462
43	232.2	46.4	544	9	A1380421	A1380421	tf96d02.x
44	225.4	45.9	871	10	BI147007	BI147007	602913462
45	225.4	45.2	508	9	AM540879	AM540879	C0138C05-

ALIGNMENTS

RESULT 1
BF037074
LOCUS 601456938BF1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3860504 5',
DEFINITION mRNA sequence.
BF037074
BF037074.1 GI:10745303

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/.
Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCPD/DTP
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LAM9595 row: 1 column: 09
High quality sequence stop: 738.
Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3860504"
/clone_id="NIH_MGC_66"
/tissue_type="adrenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: PCMV-SPORE; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 233 a 163 c 186 g 179 t

ORIGIN

Query Match 100.0%; Score 500; DB 10; Length 761;
 Best Local Similarity 100.0%; Pred. No. 4.6e-117;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

01 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCCGCGCCATCATGCTCTTCCCAA 60
 |||||||
 Db 69 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCCGCGCCATCATGCTCTTCCCAA 128
 |||||||
 02 GGGCGCGGAGCG 120
 |||||||
 Db 129 GGGCGCGGAGCG 188
 |||||||
 03 GTCAGATTGACCTGTGGAGAGATGTTGAAGCTTTGTTAGAGAGAGAGAGAGAGATG 248
 |||||||
 Db 189 GTCAGATTGACCTGTGGAGAGATGTTGAAGCTTTGTTAGAGAGAGAGAGAGATG 248
 |||||||
 04 CTTTCGCCCTTTTCAGGAAGTGGATTAAGAGAGATTAAGAGAGATTAAGAGAGATG 308
 |||||||
 Db 249 CTTTCGCCCTTTTCAGGAAGTGGATTAAGAGAGATTAAGAGAGATTAAGAGAGATG 308
 |||||||
 05 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 300
 |||||||
 Db 241 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 300
 |||||||
 06 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 368
 |||||||
 Db 309 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 368
 |||||||
 07 ACCAAGATCATATGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGAT 360
 |||||||
 Db 369 ACCAAGATCATATGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGAT 428
 |||||||
 08 ACATCAAGACAGCTCATTAAGCAATGTTTAAACATCACCAGAACGGAAGACAG 488
 |||||||
 Db 429 ACATCAAGACAGCTCATTAAGCAATGTTTAAACATCACCAGAACGGAAGACAG 488
 |||||||
 09 CTTGCGTACCAATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 480
 |||||||
 Db 421 CTTGCGTACCAATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 480
 |||||||
 10 CTTGCGTACCAATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 548
 |||||||
 Db 489 CTTGCGTACCAATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 548
 |||||||
 11 AATGTTATCTGATCCAGTG 500
 |||||||
 Db 481 AATGTTATCTGATCCAGTG 500
 |||||||
 12 AATGTTATCTGATCCAGTG 568
 |||||||
 Db 549 AATGTTATCTGATCCAGTG 568
 |||||||

RESULT 2

AU121047 844 bp mRNA linear EST 19-OCT-2000

LOCUS AU121047 HEMBI Homo sapiens cDNA clone HEMBI1001966 5', mRNA

DEFINITION AU121047 HEMBI Homo sapiens cDNA clone HEMBI1001966 5', mRNA

ACCESSION AU121047 GI:10936282

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 844)

AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, D., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and

Isogai, T.

TITLE HRI human cDNA project

JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3951

Fax: 81-438-52-3952

Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix

Research Institute; cDNA library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and

Helix Research Institute.

Location/Qualifiers

source

1. 844

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HEMBI1001966"

/clone_lib="HEMBI1"

/tissue_type="whole embryo, mainly body"

/dev_stage="embryo, 10 weeks"

/note="Vector: pME18FL3"

BASE COUNT 252 a 179 c 205 g 204 t 4 others

ORIGIN

Query Match 100.0%; Score 500; DB 9; Length 844;
 Best Local Similarity 100.0%; Pred. No. 4.8e-117;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

01 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCCGCGCCATCATGCTCTTCCCAA 60
 |||||||
 Db 106 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCCGCGCCATCATGCTCTTCCCAA 165
 |||||||
 02 GGGCGCGGAGCG 120
 |||||||
 Db 61 GGGCGCGGAGCG 225
 |||||||
 03 GTCAGATTGACCTGTGGAGAGATGTTGAAGCTTTGTTAGAGAGAGAGAGAGATG 180
 |||||||
 Db 121 GTCAGATTGACCTGTGGAGAGATGTTGAAGCTTTGTTAGAGAGAGAGAGAGATG 180
 |||||||
 04 CTTTCGCCCTTTTCAGGAAGTGGATTAAGAGATTAAGAGATTAAGAGAGATTAAGAG 285
 |||||||
 Db 226 CTTTCGCCCTTTTCAGGAAGTGGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 285
 |||||||
 05 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 300
 |||||||
 Db 241 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 300
 |||||||
 06 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 405
 |||||||
 Db 346 AAGATTGCGAGTTGGAAGATTTCCCATGCTGAAACACTTGTATCCCAAGACTGGAGA 405
 |||||||
 07 ACCAAGATCATATGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGAT 360
 |||||||
 Db 301 ACCAAGATCATATGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGATTTGGAGAT 360
 |||||||
 08 ACATCAAGACAGCTCATTAAGCAATGTTTAAACATCACCAGAACGGAAGACAG 465
 |||||||
 Db 406 ACATCAAGACAGCTCATTAAGCAATGTTTAAACATCACCAGAACGGAAGACAG 465
 |||||||
 09 CTTGCGTACCAATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 420
 |||||||
 Db 361 CTTGCGTACCAATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 420
 |||||||
 10 AATGTTATCTGATCCAGTG 500
 |||||||
 Db 466 AATGTTATCTGATCCAGTG 500
 |||||||
 11 AATGTTATCTGATCCAGTG 525
 |||||||
 Db 421 AATGTTATCTGATCCAGTG 525
 |||||||
 12 AATGTTATCTGATCCAGTG 585
 |||||||
 Db 526 AATGTTATCTGATCCAGTG 585
 |||||||
 13 AATGTTATCTGATCCAGTG 605
 |||||||
 Db 586 AATGTTATCTGATCCAGTG 605
 |||||||

RESULT 3

BC012376 920 bp mRNA linear HTC 20-AUG-2001

LOCUS BC012376 Homo sapiens, clone IMAGE:3860504, mRNA.

DEFINITION BC012376 Homo sapiens, clone IMAGE:3860504, mRNA.

ACCESSION BC012376 GI:15214504

KEYWORDS HTC.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 920)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (15-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgabds-r@mail.nih.gov

REMARK COMMENT

Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalob@bcm.tmc.edu.
 Villalob, D.K., Luna, R.A., Hale, S.M., Huijck, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
 Series: IRAX Plate: 21 Row: e Column: 16
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Genomescan gene prediction, similarity but not identity
 to protein

This clone has the following problem: incomplete processing.

FEATURES

source

Location/Qualifiers
 1..920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3860504"
 /tissue_type="Ovary, adenocarcinoma"
 /clone_lib="NIH_MGC_66"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 BASE COUNT 304 a 180 c 222 g 214 t
 ORIGIN

Query Match 100.0%; Score 500; DB 11; Length 920;
 Best Local Similarity 100.0%; Pred. No. 5e-117;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCGGGCCATCATGCTCTTGGCCAA 60
 |||||
 89 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCGGGCCATCATGCTCTTGGCCAA 148
 |||||
 61 GCGCGCGGAGCGCGGGGCGACCGCCATCACCAGCTCTGACCTGCTCTTACACTCT 120
 |||||
 149 GCGCGCGGAGCGCGGGGCGACCGCCATCACCAGCTCTGACCTGCTCTTACACTCT 208
 |||||
 121 GTCAATTTGAACCTGTGGGAAGATGTTTGAAGCTTTTGAAGGAGGAAACAAATG 180
 |||||
 209 GTCAATTTGAACCTGTGGGAAGATGTTTGAAGCTTTTGAAGGAGGAAACAAATG 268
 |||||
 181 CTCTGCTCTTTTTCAGAACTGAGAGATAGAAAGATTATCCAGGAATCAGAAATG 240
 |||||
 269 CTCTGCTCTTTTTCAGAACTGAGAGATAGAAAGATTATCCAGGAATCAGAAATG 328
 |||||
 241 AAGAATTTGAGTTGGGAAGGTTTCCATGCTGAAAACATTGATCCCAAGACTGGAAGA 300
 |||||
 329 AAGAATTTGAGTTGGGAAGGTTTCCATGCTGAAAACATTGATCCCAAGACTGGAAGA 388
 |||||
 301 ACCAAGATCATTTAGAGTCTTGGACTTGGCCATGTGAGATACAAAGGCTACACAGAGAC 360
 |||||
 389 ACCAAGATCATTTAGAGTCTTGGACTTGGCCATGTGAGATACAAAGGCTACACAGAGAC 448
 |||||
 361 AGATTCAAAGAGCTCATTAAGCAATGTTTAAACATCACCCAGCAACAAAGGAGAG 420
 |||||
 449 AGATTCAAAGAGCTCATTAAGCAATGTTTAAACATCACCCAGCAACAAAGGAGAG 508
 |||||
 421 CTGTGTGAACCAATTAAGGAGATTAATGACTACTCTTGCATTAAGAGCTTATG 480
 |||||
 509 CTGTGTGAACCAATTAAGGAGATTAATGACTACTCTTGCATTAAGAGCTTATG 568
 |||||
 481 AAATGTATTCATCCAGTG 500
 |||||
 569 AAATGTATTCATCCAGTG 588

RESULT 4

BG720251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

602692311f1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4824555 5',
 mRNA sequence.
 BG720251
 BG720251.1 GI:13999438
 EST.
 human.
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 666)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraiki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: LLAM10736 Row: e Column: 04
 high quality sequence stop: 661.
 Location/Qualifiers
 1..666
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4824555"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptPR (modified
 pluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtagag
); Oligo-dT primed using primer 5'-TTTATTTTATTTTATTTT-3',
 size-selected for average insert size 2.2 kb and
 normalized to 10^5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NHGRI/NHRI, National
 Institutes of Health). Note: this is a NIH-MGC Library."

FEATURES

source

BASE COUNT 198 a 155 c 173 g 140 t
 ORIGIN

Query Match 97.8%; Score 489; DB 10; Length 666;
 Best Local Similarity 99.8%; Pred. No. 2.8e-114;
 Matches 500; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCGGGCCATCATGCTCTTGGCCAA 60
 |||||
 94 GGGACGTGAGCCGCTGCGCCACCGGGCTAGACCGGGCCATCATGCTCTTGGCCAA 153
 |||||
 61 GCGCGCGGAGCGCGGGGCGACCGCCATCACCAGCTCTGACCTGCTCTTACACTCT 120
 |||||
 154 GCGCGCGGAGCGCGGGGCGACCGCCATCACCAGCTCTGACCTGCTCTTACACTCT 213
 |||||
 121 GTCAATTTGAACCTGTGGGAAGATGTTTGAAGCTTTTGAAGGAGGAAACAAATG 180
 |||||
 214 GTCAATTTGAACCTGTGGGAAGATGTTTGAAGCTTTTGAAGGAGGAAACAAATG 273
 |||||
 181 CTCTGCTCTTTTTCAGAACTGAGAGATAGAAAGATTATCCAGGAATCAGAAATG 240
 |||||
 274 CTCTGCTCTTTTTCAGAACTGAGAGATAGAAAGATTATCCAGGAATCAGAAATG 333
 |||||
 241 AAGAATTTGAGTTGGGAAGGTTTCCATGCTGAAAACATTGATCCCAAGACTGGAAGA 300
 |||||
 334 AAGAATTTGAGTTGGGAAGGTTTCCATGCTGAAAACATTGATCCCAAGACTGGAAGA 393
 |||||
 301 ACCAAGATCATTTAGAGTCTTGGACTTGGCCATGTGAGATACAAAGGCTACACAGAGAC 360
 |||||

Db 394 ACCAAGATCATTTGAGTCTTGGAGCTTGGCCATGTGAGATACAGGCTACACAGAGAC 453

QY 361 AGATCAAGCAGCTCTTAAGCAATGTTTAAACATCACCACCAACAAGGAAAGAG 420

Db 454 AGATCAAGCAGCTCTTAAGCAATGTTTAAACATCACCACCAACAAGGAAAGAG 513

QY 421 CTGTGTAACCAATTAAGAGAGATATGACTTACTTGCAT-ACCTAAGCTTAT 479

Db 514 CTGTGTAACCAATTAAGAGAGATATGACTTACTTGCAT-ACCTAAGCTTAT 573

QY 480 GAAATGTTATCTGATCCAGTG 500

Db 574 GAAATGTTATCTGATCCAGTG 594

RESULT 5
AL548837 780 bp mRNA linear EST 16-FEB-2001
LOCUS AL548837 LFI.NFL006.PL2 Homo sapiens cDNA clone CS0D1043Y07 5
DEFINITION prime, mRNA sequence.
ACCESSION AL548837
VERSION AL548837.1 GI:12884235
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 780)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1043Y07"
/clone.lib="LFI.NFL006.PL2"
/issue.type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 232 a 179 c 194 g 173 t 2 others

ORIGIN

Query Match 97.8%; Score 489; DB 9; Length 780;
Best Local Similarity 99.8%; Pred. No. 3e-114;
Matches 500; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGGACGTGAGCCCGTGGCCACCGGGCTAAGACCCGGCCCATCATGCTTCTGCCAA 60

Db 129 GGGACGTGAGCCCGTGGCCACCGGGCTAAGACCCGGCCCATCATGCTTCTGCCAA 188

QY 61 GCGGGCGGG-AGCGCGGGGGGACCGGCATCACCAAGCTTCCCTCTACACTC 119

Db 189 GCGGGCGGGAGCGGGGGGACCGGCATCACCAAGCTTCCCTCTACACTC 248

QY 120 TGTCAAGTTAAGCTGTGGAGAGATGTTGAAGCTTTGTTAAGAGAGAAAGAAAT 179

Db 249 TGTCAAGTTAAGCTGTGGAGAGATGTTGAAGCTTTGTTAAGAGAGAAAGAAAT 308

QY 180 GCTTCTCCCTCTTTCGAAAGTGAAGATTAAGAGAGTTCCAGAGATCACAAGAT 239

Db 309 GCTTCTCCCTCTTTCGAAAGTGAAGATTAAGAGAGTTTCCAGAGATCAGAAT 368

QY 240 GAGAAATTCAGTCTGAGAGAGTTCATGCTGAAGAAACATGTTCCCAAGAGTGAAG 299

Db 369 GAGAAATTCAGTCTGAGAGAGTTCATGCTGAAGAAACATGTTCCCAAGAGTGAAG 428

QY 300 AACCAAGATCATTTATGACAGTCTTGGAGTGGCCATGAGATACAAGCTACACAGAGA 359

Db 429 AACCAAGATCATTTATGACAGTCTTGGAGTGGCCATGAGATACAAGCTACACAGAGA 488

QY 360 CAGATCAAGCAGCTCATTAAGCAATGTTTAAACATCACCACCAACAAGGAAAGCA 419

Db 489 CAGATCAAGCAGCTCATTAAGCAATGTTTAAACATCACCACCAACAAGGAAAGCA 548

QY 420 GCTGTGAACCAATTAAGAGAGATATGACTTACTTGCAT-ACCTAAGCTTAT 479

Db 549 GCTGTGAACCAATTAAGAGAGATATGACTTACTTGCAT-ACCTAAGCTTAT 608

QY 480 GAAATGTTATCTGATCCAGTG 500

Db 609 GAAATGTTATCTGATCCAGTG 629

RESULT 6
B1460193 689 bp mRNA linear EST 21-ANG-2001
LOCUS B1460193 NTH_MGC_97 Homo sapiens cDNA clone IMAGE:5267854 5',
DEFINITION mRNA sequence.
ACCESSION B1460193
VERSION B1460193.1 GI:15250849
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rt@mail.nih.gov
Tissue Procurement: Miklos Palkevits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
size-selected for average insert size 2.2 kb and
normalized to 10⁶ 5'. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 209 a 156 c 176 g 148 t

ORIGIN

Query Match 97.6%; Score 488; DB 10; Length 689;
Best Local Similarity 99.8%; Pred. No. 5.2e-114;
Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGGACGTGAGCCCGCTGCGCCACCGGCTAGACCCGGCGGCATCATGCTTCTGCCAA 60
 DB 72 GGGACGTGAGCCCGCTGCGCCACCGGCTAGACCCGGCGGCATCATGCTTCTGCCAA 131
 QY 61 GCGCCGGGACCGCGCGGCGACCCGATACCCGCTGAGCTTGGCTTCTGACCT 120
 DB 132 GCGCCGGGACCGCGGCGGCGACCCGATACCCGCTGAGCTTGGCTTCTGACCT 191
 QY 121 GTCAGTTGACCTGTGGAGAGATGCTTTGAAGCTTTGTTAAGAGAGACAAGATG 180
 DB 192 GTCAGTTGACCTGTGGAGAGATGCTTTGAAGCTTTGTTAAGAGAGACAAGATG 250
 QY 181 CTCTGCTCTCTTTTCAGGAATGAGATTAAGAAAGATTATCCGAGGAATCAGAATG 240
 DB 251 CTCTGCTCTCTTTTCAGGAATGAGATTAAGAAAGATTATCCGAGGAATCAGAATG 310
 QY 241 AAGAATTGCGAGTTGGAAGATTTCCTCATGCTGAACACTTGCCTCAAGACTGGAAGA 300
 DB 311 AAGAATTGCGAGTTGGAAGATTTCCTCATGCTGAACACTTGCCTCAAGACTGGAAGA 370
 QY 301 ACCAAGATCATTTAGTTCGAGCTTGGCATGTGAGATCAAGAGCTACACAGAGAC 360
 DB 371 ACCAAGATCATTTAGTTCGAGCTTGGCATGTGAGATCAAGAGCTACACAGAGAC 430
 QY 361 AGATCAAGAGAGCTCATTAAGCAATGTTTAAACATCAACCCAGCAAGAGAGAGAG 420
 DB 431 AGATCAAGAGAGCTCATTAAGCAATGTTTAAACATCAACCCAGCAAGAGAGAGAG 490
 QY 421 CTGCTGAACCAATTAAGAGAGATTAATGACTTCTTCACTTGCATTAAGAGCTATG 480
 DB 491 CTGCTGAACCAATTAAGAGAGATTAATGACTTCTTCACTTGCATTAAGAGCTATG 550
 QY 481 AATGTTATCTGATCCAGTG 500
 DB 551 AATGTTATCTGATCCAGTG 570

RESULT 7 937 bp mRNA linear EST 12-MAR-2001
 LOCUS BG396661
 DEFINITION 602459730F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4582038 5',
 mRNA sequence.
 ACCESSION BG396661
 VERSION BG396661.1 GI:13290109
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 937)
 AUTHORS NIH-MGC, <http://imgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9abps-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Lung Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: imgc.llnl.gov
 Plate: LCM1303 row: 1 column: 07
 High quality sequence stop: 758.
 location/Qualifiers
 1. 937
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4582038"
 /clone_id="NIH_MGC_16"
 /issue_type="retrotransposon"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:

BASE COUNT 280 a 192 c 244 g 221 t
 ORIGIN
 Query Match 97.6%; Score 488; DB 10; Length 937;
 Best Local Similarity 99.8%; Pred. No. 5,8e-114;
 Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 GGGACGTGAGCCCGCTGCGCCACCGGCTAGACCCGGCGGCATCATGCTTCTGCCAA 60
 DB 86 GGGACGTGAGCCCGCTGCGCCACCGGCTAGACCCGGCGGCATCATGCTTCTGCCAA 145
 QY 61 GCGCCGGGACCGCGCGGCGACCCGATACCCGCTGAGCTTGGCTTCTGACCT 120
 DB 146 GCGCCGGGACCGCGGCGGCGACCCGATACCCGCTGAGCTTGGCTTCTGACCT 205
 QY 121 GTCAGTTGACCTGTGGAGAGATGCTTTGAAGCTTTGTTAAGAGAGACAAGATG 180
 DB 206 GTCAGTTGACCTGTGGAGAGATGCTTTGAAGCTTTGTTAAGAGAGACAAGATG 265
 QY 181 CTCTGCTCTCTTTTCAGGAATGAGATTAAGAAAGATTATCCGAGGAATCAGAATG 240
 DB 266 CTCTGCTCTCTTTTCAGGAATGAGATTAAGAAAGATTATCCGAGGAATCAGAATG 325
 QY 241 AAGAATTGCGAGTTGGAAGATTTCCTCATGCTGAACACTTGCCTCAAGACTGGAAGA 300
 DB 326 AAGAATTGCGAGTTGGAAGATTTCCTCATGCTGAACACTTGCCTCAAGACTGGAAGA 385
 QY 301 ACCAAGATCATTTAGTTCGAGCTTGGCATGTGAGATCAAGAGCTACACAGAGAC 360
 DB 386 ACCAAGATCATTTAGTTCGAGCTTGGCATGTGAGATCAAGAGCTACACAGAGAC 444
 QY 421 CTGCTGAACCAATTAAGAGAGATTAATGACTTCTTCACTTGCATTAAGAGCTATG 480
 DB 505 CTGCTGAACCAATTAAGAGAGATTAATGACTTCTTCACTTGCATTAAGAGCTATG 564
 QY 481 AATGTTATCTGATCCAGTG 500
 DB 565 AATGTTATCTGATCCAGTG 584

RESULT 8
 LOCUS AL517760 955 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL517760 LRL_NFL011.NBC1 Homo sapiens cDNA clone CS00A004YR24 5
 prime, mRNA sequence.
 ACCESSION AL517760
 VERSION AL517760.1 GI:12781253
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 955)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 location/Qualifiers
 1. 955

```

/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="CS00A004YE24"
/clone.lib="LRI_NF011_NBC1"
/sex="male"
/tissue="neuroblastoma cells"
/lab.host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library is not normalized, but is the control for
the normalized libraries. Library was constructed by Life
Technologies. Contact: Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax: (1) 301 610 8371 Email:
f.liang@lifestech.com URL:
http://fulllength.invitrogen.com"
BASE COUNT      328 a      174 c      222 g      221 t      10 others
ORIGIN
Query Match      97.3%; Score 486.6; DB 9; Length 955;
Best Local Similarity 98.6%; Pred. No. 1,3e-113;
Matches 494; Conservative 6; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGGACGTGAGCCGCTGCGCCACCGGCTAGACCGCGCCATCATGCTGCTGCCAA 60
    |||||
DB 49 GGGACGTGAGCCGCTGCGCCACCGGCTAGACCGCGCCATCATGCTGCTGCCAA 108
    |||||

QY 61 GCGCGCGCGG-ACGGCGGGGGGCGACCGCCATCCAGCGCTGCTGCTGACATC 119
    |||||
DB 109 GCGCGCGGGAGCGCGGGGGGCGACCGCCATCCAGCGCTGCTGCTGACATC 168
    |||||

QY 120 TGTCAAGTTGAACCTGTGGAGAGATGTTGAAGCTTTTGTAGAGGAGAAACAAT 179
    |||||
DB 169 TGTCAAGTTGAACCTGTGGAGAGATGTTGAAGCTTTTGTAGAGGAGAAACAAT 228
    |||||

QY 180 GCTTTCGCTCTTTTCAGAACTGAGATGAGAAAGATTAACGAGAAATCAAAAT 239
    |||||
DB 229 GCTTTCGCTCTTTTCAGAACTGAGATGAGAAAGATTAACGAGAAATCAAAAT 288
    |||||

QY 240 GAAGAATTGCAATGGAGAGATTTCCATGCTGTAAGAACTGTGACCAAGCTGGA 299
    |||||
DB 289 GAAGAATTGCAATGGAGAGATTTCCATGCTGTAAGAACTGTGACCAAGCTGGA 348
    |||||

QY 300 AACCAAGATCATTAATGAGATTTGAGACTTGGCCATGTGATACAAAGCTACAGAGA 359
    |||||
DB 349 AACCAAGATCATTAATGAGATTTGAGACTTGGCCATGTGATACAAAGCTACAGAGA 408
    |||||

QY 360 CAGATCAAAAGCAGCTCATAAAGCAATGTTTAAACATCACCCAGCAAAAGGAAACA 419
    |||||
DB 409 CAGATCAAAAGCAGCTCATAAAGCAATGTTTAAACATCACCCAGCAAAAGGAAACA 468
    |||||

QY 420 GCTGTGGAACCAATAAAGAGAGATTAATGACTTCACTTGCATTAACCTTAT 479
    |||||
DB 469 GCTGTGGAACCAATAAAGAGAGATTAATGACTTCACTTGCATTAACCTTAT 528
    |||||

QY 480 GAAATGTATCTGATCCAGTG 500
    |||||
DB 529 GAAATGTATCTGATCCAGTG 549
    |||||

RESULT 9
LOCUS      BG718443      829 bp      mRNA      linear      EST 08-MAY-2001
DEFINITION 602696473P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828339 5',
ACCESSION  BG718443
VERSION    BG718443.1 GI:13997630
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE    1 (bases 1 to 829)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs.fda.nih.gov
            Tissue Procurement: Miklos Palkovics, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NIGRI), Shitaki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            plate: LLM10746 row: D column: 20
            High quality sequence stop: 822.
            Location/Qualifiers
                1..829
                /organism="Homo sapiens"
                /db.xref="taxon:9606"
                /clone="IMAGE:4828339"
                /clone.lib="NIH_MGC_97"
                /lab.host="DH10B"
                /note="Organ: testis; Vector: pBluescript (modified
                bluescript KS+); Site: 1: BamHI; Site: 2: SalI-XhoI (gtcgag
                ); Oligo-dT primed using primer 5'-TTCCTTCTTTTCTTCTT-3',
                size-selected for average insert size 2.2 kb and
                normalized to ROP 5. This is a primary library enriched
                for full-length clones and constructed using the
                cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIGRI/NHRI, National
                Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      263 a      165 c      201 g      199 t      1 others
ORIGIN
Query Match      97.1%; Score 485.4; DB 10; Length 829;
Best Local Similarity 99.4%; Pred. No. 2,3e-113;
Matches 497; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GGGACGTGAGCCGCTGCGCCACCGGCTAGACCGCGCCATCATGCTGCTGCCAA 60
    |||||
DB 58 GGGACGTGAGCCGCTGCGCCACCGGCTAGACCGCGCCATCATGCTGCTGCCAA 117
    |||||

QY 61 GCGCGCGGAGCGCGGGGCGACCGCCATCCAGCGCTGCTGCTGACATCT 120
    |||||
DB 118 GCGCGCGGAGCGCGGGGCGACCGCCATCCAGCGCTGCTGCTGACATCT 176
    |||||

QY 121 GTCAGTTGAACCTGTGGAGAGATGTTGAAGCTTTTGTAGAGGAGAAACAAT 180
    |||||
DB 177 GTCAGTTGAACCTGTGGAGAGATGTTGAAGCTTTTGTAGAGGAGAAACAAT 236
    |||||

QY 181 CTTCTGCTCTTTTCAGAACTGAGATGAGAAAGATTAACGAGAAATCAAAAT 240
    |||||
DB 237 CTTCTGCTCTTTTCAGAACTGAGATGAGAAAGATTAACGAGAAATCAAAAT 296
    |||||

QY 241 AAGAAATGCAAGTTGGAAGAGTTTCCATGCTGAAGAACATCTGATCCCAAGACTGGAAGA 300
    |||||
DB 297 AAGAAATGCAAGTTGGAAGAGTTTCCATGCTGAAGAACATCTGATCCCAAGACTGGAAGA 356
    |||||

QY 301 ACCAAGATCATTAATGAGATTTGAGACTTGGCCATGTGATACAAAGCTACAGAGA 360
    |||||
DB 357 ACCAAGATCATTAATGAGATTTGAGACTTGGCCATGTGATACAAAGCTACAGAGA 416
    |||||

QY 361 AGATCAAGAGAGCTCATTAAGCAATGTTTAAACATCCAGCAAAAGGAAAGAG 420
    |||||
DB 417 AGATCAAGAGAGCTCATTAAGCAATGTTTAAACATCCAGCAAAAGGAAAGAG 476
    |||||

QY 421 CTGCTGAGCCATTAAGAGAGAGATTAATGACTTCACTTGCATTAACCTTAT 480
    |||||
DB 477 CTGCTGAGCCATTAAGAGAGAGATTAATGACTTCACTTGCATTAACCTTAT 536
    |||||

QY 481 AATGTATCTGATCCAGTG 500
    |||||

```

```

Db      537 AAATGTTATCTGATCCAGTG 556
|||||
RESULT 10
AV655598
LOCUS      AV655598      665 bp      mRNA      linear      EST 15-JAN-2002
DEFINITION AV655598      GLC Homo sapiens cDNA clone G1CE1E03 3', mRNA sequence.
ACCESSION  AV655598
VERSION     AV655598.1  GI:9876612
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 665)
AUTHORS     Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
             Xiao,H., Qu,D., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
             Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
             Hu,G., Gu,J., Chen,Z., and Han,Z.
             Insight into hepatocellular carcinogenesis at transcriptome level
             by comparing gene expression profiles of hepatocellular carcinoma
             with those of corresponding noncancerous liver
             Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

JOURNAL
MEDLINE
COMMENT
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source      Location/Qualifiers
             1..665
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="G1CE1E03"
             /clone_lib="GLC"
             /tissue_type="corresponding non cancerous liver tissue"
             /dev_stage="Adult"
             /lab_host="SOLR"
             /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
             XhoI"

BASE COUNT      189 a      159 c      172 g      145 t
ORIGIN
Query Match      96.6%; Score 483.2; DB 9; Length 665;
Best Local Similarity 99.2%; Pred. No. 8.5e-113;
Matches 496; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY      1 GGGACGTGAGCGGCTGCGCCACCGGCTAGACCGGCGCCATCGTCTTGCCCA 60
      |||||||
Db      113 GGGACGTGAGCGGCTGCGCCACCGGCTAGACCGGCGCCATCGTCTTGCCCA 172

QY      61 GCGCCGCGAGCGCGGCGGACCGGCATCACCGCTCTGACCTCTCTACACTCT 120
      |||||||
Db      173 GCGCCGCGAGCGCGGCGGACCGGCATCACCGCTCTGACCTCTCTACACTCT 232

QY      121 GTCAATTTGACCTGTTGAGAGATGTTGAGCTTTTGTAGAGGAGAAACAGAAATG 180
      |||||||
Db      233 GTCAATTTGACCTGTTGAGAGATGTTGAGCTTTTGTAGAGGAGAAACAGAAATG 292

QY      181 CTTCTGCTCTTTTTCAGAGACTGAGAGATAGAGAAAGTTATCCAGAGATCAGAGATG 240
      |||||||
Db      293 CTTCTGCTCTTTTTCAGAGACTGAGAGATAGAGAAAGTTATCCAGAGATCAGAGATG 352

QY      241 AAGAATTGAGTTGAGAGATTTCCATGCTGAAAAACACTTGATCCCAAGACTGGAAGA 300
      |||||||
Db      353 AAGAATTGAGTTGAGAGATTTCCATGCTGAAAAACACTTGATCCCAAGACTGGAAGA 412

QY      301 ACGAAGATGATTTGAGAGTTCTTGACCTTGCCATGTGAGATACAGAGCTACAGAGAC 360
      |||||||

```

```

Db      413 ACCAAGATCATTTATGAGTTCTTGACCTTGCCATGTGAGATACAGAGCTACAGAGAC 472
QY      361 AGATCAAGACGCTCATTAAGCAATGTTTAAACATCACCCAGACAGAGAGAGAG 420
      |||||||
Db      473 AGATCAAGACGCTCATTAAGCAATGTTTAAACATCACCCAGACAGAGAGAGAG 532

QY      421 CTGGTGAACCATTAAGAGAGATATGACTACTTCTGCACTTAAGCTTATG 480
      |||||||
Db      533 CTGGTGAACCATTAAGAGAGATATGACTACTTCTGCACTTAAGCTTATG 591

QY      481 AAATGTTATCTGATCCAGTG 500
      |||||
Db      592 AAATGTTATCTGATCCAGTG 611

RESULT 11
BG165721
LOCUS      BG165721      859 bp      mRNA      linear      EST 06-FEB-2001
DEFINITION 602344558F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4454366 5',
             mRNA sequence.
ACCESSION  BG165721
VERSION     BG165721.1  GI:12672424
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 859)
AUTHORS     NIH-MGC http://mgi.mcl.nih.gov/.
             National Institutes of Health, Mammalian Gene Collection (MGC)
             Unpublished (1999)
             Contact: Robert Strausberg, Ph.D.
             Email: cgrabs@remail.nih.gov
             Tissue Procurement: ARCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LNL at:
             http://image.llnl.gov
             Plate: LLM10246 row: d column: 15
             High quality sequence stop: 743.

FEATURES
source      Location/Qualifiers
             1..859
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:4454366"
             /clone_lib="NIH_MGC_89"
             /tissue_type="hypernephroma, cell line"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: Kidney; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
             Average insert size 1.3 kb. Library enriched for
             full-length clones and constructed by Life Technologies.
             Note: this is a NIH_MGC library."

BASE COUNT      287 a      157 c      208 g      207 t
ORIGIN
Query Match      95.2%; Score 476; DB 10; Length 859;
Best Local Similarity 99.6%; Pred. No. 6.5e-111;
Matches 498; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY      1 GGGACGTGAGCGGCTGCGCCACCGGCTAGACCGGCGCCATCATGCTTTCGCCAA 60
      |||||||
Db      1 GGGACGTGAGCGGCTGCGCCACCGGCTAGACCGGCGCCATCATGCTTTCGCCAA 60

QY      61 GCGCCGCGAGCGCGGCGGACCGGCATCACCGCTCTGACCTCTGCTTACACTCT 120
      |||||||
Db      61 GCGCCGCGAGCGCGGCGGACCGGCATCACCGCTCTGACCTCTGCTTACACTCT 120

QY      121 GTCAATTTGACCTGTTGAGAGATGTTGAGCTTTTGTAGAGAGAGAAACAGAAATG 180
      |||||||
Db      121 GTCAATTTGACCTGTTGAGAGATGTTGAGCTTTTGTAGAGAGAGAAACAGAAATG 179

```

OY 181 CTTTCGCTCTTTTTCAGAACTGAGATTAAGAAAGAGTTATCCGAGGATCAGAAAGATG 240
 |||||||
 DB 180 CTTTCGCTCTTTTTCAGAACTGAGATTAAGAAAGAGTTATCCGAGGATCAGAAAGATG 239
 |||||||
 OY 241 AAGAAATGAGTTGGAGAGTTTCCATGCTGAAACAACTTATCCCAAGACAGTGAAGA 300
 |||||||
 DB 240 AAGAA-TGCAAGTTGGAGAGTTTCCATGCTGAAACAACTTATCCCAAGACAGTGAAGA 298
 |||||||
 OY 301 ACCAAGATCATTAATGAGTTCTTGGAGTTGGCCATGTAGATACAAAGGCTACACAGAGAC 360
 |||||||
 DB 299 ACCAAGATCATTAATGAGTTCTTGGAGTTGGCCATGTAGATACAAAGGCTACACAGAGAC 358
 |||||||
 OY 361 AGATCAAGACAGCTGATTAAGCAATGTTTAAACATCACCCAGACAAAGAGAGAGAG 420
 |||||||
 DB 359 AGATCAAGACAGCTGATTAAGCAATGTTTAAACATCACCCAGACAAAGAGAGAGAG 418
 |||||||
 OY 421 CTTGTGAACCATTAAGAGAGATTAATGACTTACTTGCATTAACCTAAAGCTTATG 480
 |||||||
 DB 419 CTGTGTGAACCATTAAGAGAGATTAATGACTTACTTGCATTAACCTAAAGCTTATG 478
 |||||||
 OY 481 AATGTATTCGATCCAGTG 500
 |||||||
 DB 479 AATGTATTCGATCCAGTG 498

RESULT 12
 A1364977/c 735 bp mRNA linear EST 16-FEB-1999
 LOCUS A1364977

DEFINITION g241h10.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2029507 3' similar to TR:099543 099543 M-PHASE PHOSPHOPROTEIN 11 ; mRNA sequence.

ACCESSION A1364977
 VERSION A1364977
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 735)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 TITLE Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail@nci.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 572 Std Error: 0.00
 Seq primer: 400P from Gibco
 High quality sequence stop: 371.

FEATURES

SOURCE

1. 735
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2029507"
 /clone_lib="NCI-CGAP_Kid11"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1323376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M.

BASE COUNT 194 a 162 c 144 g 233 t 2 others
 ORIGIN

Query Match 94.6%; Score 472.8; DB 9; Length 735;
 Best local Similarity 98.0%; Pred No. 4e-110;
 Matches 488; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

OY 3 GACGTGAGCCGCTGCGCCACCGGGGCTAGACCGGCGCCATCATGCTCTTTCGCAAGC 62
 |||||||
 DB 735 GACGTGAGNGCTGCGCCACCGGCTAGAGACCGGCGCCATCATGCTCTTTCGCAAGC 676
 |||||||
 OY 63 GCCGAGACGCGCGGCGCAACCCGATCACCGCTGACCTTGCCTTACACTGT 122
 |||||||
 DB 675 GCCGAGAGGAGGAGGCGACCCGATCACCGCTGACCTTGCCTTACACTGT 616
 |||||||
 OY 123 CAAGTTGAACCTGTGGGAAGATGTTGAAGCTTTTGTAAAGAGAAAGAAATGCT 182
 |||||||
 DB 615 CAAGTTGAACCTGTGGGAAGATGTTGAAGCTTTTGTAAAGAGAAAGAAATGCT 556
 |||||||
 OY 183 TCTGCTCTTTTCAGGAGACTGAGATTAAGAAAGATTATCCGAGAAATCAGAGATGAA 242
 |||||||
 DB 555 TCTGCTCTTTTCAGGAGACTGAGATTAAGAAAGATTATCCGAGAAATCAGAGATGAA 496
 |||||||
 OY 243 GAATTCAGTTGGAGAGATGTTCCATGCTGAAGAAACACTTGATCCCAAGACTGGAGAGAC 302
 |||||||
 DB 495 GAATTCAGTTGGAGAGATGTTCCATGCTGAAGAAACACTTGATCCCAAGACTGGAGAGAC 437
 |||||||
 OY 303 CAAGATCATTAATGAGTTCTTGGAGTTGGCCATGTAGATACAAAGGCTACACAGAGACAG 362
 |||||||
 DB 436 CAAGATCATTAATGAGTTCTTGGAGTTGGCCATGTAGATACAAAGGCTACACAGAGACAG 377
 |||||||
 OY 363 ATCAAGACAGCTCATTAAGCAATGTTTAAACATCACCCAGACAAAGAGAGAGAG 422
 |||||||
 DB 376 ATCAAGACAGCTCATTAAGCAATGTTTAAACATCACCCAGACAAAGAGAGAGAG 317
 |||||||
 OY 423 GGTGAACCATTAAGAGAGATTAATGACTTACTTGCATTAACCTAAAGCTTATG 482
 |||||||
 DB 316 GGTGAACCATTAAGAGAGATTAATGACTTACTTGCATTAACCTAAAGCTTATG 257
 |||||||
 OY 483 ATGTATTCGATCCAGTG 500
 |||||||
 DB 256 ATGTATTCGATCCAGTG 239

RESULT 13
 AM161245 579 bp mRNA linear EST 09-NOV-1999
 LOCUS AM161245

DEFINITION au70d01.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2781601 5' similar to TR:060415 060415 WUGSC:H_RG276003.1A PROTEIN ; mRNA sequence.

ACCESSION AM161245
 VERSION AM161245.1 GI:6300278
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 579)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Merra, M., Martin, J., Moore, B., Schellenberg, K., Stepec, M., Tan, F., Theisling, B., White, X., Wylie, T., Waterston, R. and Wilson, R.
 WASHU-NCI human EST Project
 Unpublished (1997)
 Contact: Wilson RK
 Other ESTs: au70d01.x1

TITLE JOURNAL
 COMMENT Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LLNL ; contact the

Db 509 GCTGTCACCAATTAAGAAGAGATTAATGACTTCACTTGCATTAAGTATCC 568
QY 480 GAAATGTTATCATC 495
Db 569 CCAAGTCATTTTATTC 584

RESULT 15
BI562033
LOCUS 603255091F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5297303 5',
DEFINITION mRNA sequence.
ACCESSION BI562033
VERSION BI562033.1 GI:15449347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 656)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM1752 row: f column: 24
high quality sequence stop: 654.
Location/Qualifiers
1. 656

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5297303"
/clone_1ib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 184 a 169 c 173 g 130 t
ORIGIN

Query Match 87.6%; Score 438.2; DB 10; Length 656;
Best Local Similarity 96.8%; Pred. No. 2.6e-101;
Matches 479; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

QY 1 GGGACGTAGACCGCGTGGCCACCGGCGGTAGACCGCGCCATCATGCTGCTTGGCAA 60
Db 152 GGGACGTAGACCGCGTGGCCACCGGCGGTAGACCGCGCCATCATGCTGCTTGGCAA 211
QY 61 GCGCCGGGAGCGCGGGGACCGCCATCAACGCGTGTGACTTGGCTTACACTCT 120
Db 212 GCGCCGGGAGCGGCGGGGACCGCCATCAACGCGTGTGACTTGGCTTACACTCT 270
QY 121 GTCAAGTTGAACCTGTGGAGAGATGTTTAAGCTTTTGAAGAGAGAAACAGAAATG 180
Db 271 GTCAAGTTGAACCTGTGGAGAGATGTTTAAGCTTTTGAAGAGAGAAACAGAAATG 329
QY 181 CTTCCTGCTTTTACAGAACTGGAGATTAAGAAAGATTATCCGAGGAATCAGAGATG 240

Db 330 CTTCCTGCTTTTACAGAACTGGAGATTAAGAAAGATTATCCGAGGAATCAGAGATG 389
QY 241 AAGATTGACGTTGGAGAGATTTCCATGCGTGAAGAACCTGATCCCAAGACTGGAAGA 300
Db 390 AAGAA-TGCAGTTGGAAAGAGTTTCCATGCGTGAAGAACCTGATCCCAAGACTGGAAGA 448
QY 301 ACCAAGATCATTAATGACGTTCTTGACTGGCCATGTGAGATACAGGCTACACAGAGAC 360
Db 449 ACCAAGATCATTAATGACGTTCTTGACTGGCCATGTGAGATACAGGCTACACAGAGAC 508
QY 361 AGATCAAGCAGCTCATTAAGCAATGTTTAAACATCACCCAGACAAACGGAAGACAG 420
Db 509 AGATCAAGCAGCTCATTAAGCAATGTTTAAACATCACCCAGACAAACGGAAGACAG 568
QY 421 CTGCTGAACCAATTAAGAGAGATTAATGACTTCACTTGCATTAACCTTAATG 480
Db 569 CTGCTGAACCAATTAAGAGAGATTAATGACTTCACTTGCATTAACCTTAATG 628
QY 481 AAATGTTATCATC 495
Db 629 CAGGTGATTTTATTC 643

Search completed: October 10, 2002, 20:27:42
Job time : 702.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 84.933 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-8_COPY_1_500
Perfect score: 500
Sequence: 1 gggacgtgagcgcgtgcgc.....aatgtatctatcgcagt 500

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
- 23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001C.DAT:*
- 24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	876	21	AAC98111
2	500	100.0	1845	24	ABK09776
3	500	100.0	1860	22	ABD11115
4	80.2	16.0	2225	23	ABL16647
5	60.4	12.1	1876	22	AAH29904
6	51.8	10.4	1802	22	AAH29775
7	51.8	10.4	3083	18	AAH90969
8	51.8	10.4	3083	20	AA211218
9	42.4	8.5	4791	23	ABL16646

C	10	42.4	8.5	5071	23	ABL16648
	11	42.4	8.5	580073	18	AAH58840
	12	41.2	8.2	1039	20	AAH61812
	13	41.2	8.2	1125	20	AAH61811
C	14	41.2	8.2	35515	20	AAH20252
	15	40.8	8.2	3399	17	AAH05868
	16	40.2	8.0	2700	22	AAH17193
	17	40.2	8.0	65140	22	AAH17184
	18	40.2	8.0	125401	22	AAH17185
	19	38.4	7.7	843	23	AAH90709
	20	36.8	7.4	1686	16	AAH87587
C	21	36	7.2	1381	22	AAH32674
C	22	36	7.2	1384	22	AAH32671
C	23	36	7.2	1386	22	AAH32675
C	24	35.8	7.2	330	22	AAH64823
C	25	35.8	7.2	330	22	AAH64824
C	26	35.8	7.2	330	22	AAH64825
C	27	35.8	7.2	1305	23	AAH06649
C	28	35.8	7.2	3935	23	ABL06648
C	29	35.4	7.1	1510	17	AAH29628
C	30	35.4	7.1	2987	22	ABA08622
	31	35.4	7.1	58857	21	AAH58471
C	32	35.2	7.0	377	16	AAH90828
C	33	35.2	7.0	3579	21	AAH70099
C	34	35.2	7.0	38734	20	AAH232020
C	35	35.2	7.0	38734	22	AAH90077
	36	35	7.0	450	11	AAH03633
	37	35	7.0	1125	21	AAH56820
	38	35	7.0	4056	21	AAH70225
	39	34.8	7.0	1293	20	AAH30135
	40	34.8	7.0	910715	20	AAH20248
	41	34.6	6.9	393	22	AAH39178
	42	34.6	6.9	1924	19	AAH27403
	43	34.6	6.9	2223	16	AAH080908
	44	34.4	6.9	1372	19	AAH26236
C	45	34.4	6.9	1565	22	AAH26008

ALIGNMENTS

RESULT 1	
AAC98111	
ID	AAC98111 standard; cDNA: 876 BP.
AC	AAC98111;
XX	
DT	09-MAR-2001 (first entry)
XX	
DE	Human colon cancer antigen nucleotide sequence SEQ ID NO:121.
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	identification; cytotoxic; cardioactive; neuroprotective; vulnary;
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;
KW	neurotropic; antiinfective; antibacterial; gene therapy; wound;
KW	neural disorder; immune system disorder; muscular disorder;
KW	reproductive disorder; gastrointestinal disorder; renal disorder;
KW	infectious disease; cardiovascular disorder; ss.
OS	Homo sapiens.
XX	
PN	WO20005351-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US05883.
PR	12-MAR-1999; 99US-0124270.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	

Drosophila melanog
Myosophia melanog
B. burgdorferi ant
B. burgdorferi ant
Borrelia burgdorfe
Chicken leucocytos
ERP48 insert DNA i
Streptomyces nous
Streptomyces nous
DNA encoding leuco
Human genomic DNA
Human genomic DNA
Human immune/haema
Human immune/haema
Drosophila melanog
DNA mismatch repair
Human stromal anti
Nucleotide sequenc
Wild type hMLH1 ge
Plasmodium falcipla
Human MERT1 relate
AL021529 cDNA clon
Mycoplasma hyopneu
Mycoplasma hyopneu
Plasmodium falcipla
Streptococcus pneu
Borrelia burgdorfe
Novel human diagno
Streptococcus pneu
Plasmodium falcipla
S. pneumoniae defi
Human cDNA encodin

DR WPI: 2000-587534/55.
 XX P-PSDB: AAB53354.
 XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 XX Claim 1: Page 557: 2104pp; English.

XX AAC9791 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC and can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders, immune
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 CC
 XX Sequence 876 BP; 265 A; 186 C; 210 G; 215 T; 0 other;

Query Match 100.0%; Score 500; DB 21; Length 876;
 Best Local Similarity 100.0%; Pred. No. 1.3e-142;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGGTAGCGCGCGCCGACCGGGGTAGACCGGGGCGCATCTGCTGCTGCCAA 60
 DB 124 GGGAGGTAGCGCGCGCCGACCGGGGTAGACCGGGGCGCATCTGCTGCTGCCAA 183
 QY 61 GCGCGGAGCGCGCGGGGACCGGCATCACCCAGCTCTGACCTCTCTACACTCT 120
 DB 184 GCGCGGAGCGCGCGGGGACCGGCATCACCCAGCTCTGACCTCTCTACACTCT 243
 QY 121 GTCAAGTTGACCTGTGGAGATGTTTGAAGCTTTTGAAGAGGAGAAACAGAAATG 180
 DB 244 GTCAAGTTGACCTGTGGAGATGTTTGAAGCTTTTGAAGAGGAGAAACAGAAATG 303
 QY 181 CTTCGCTCTTTTACGAGTGGAGATAGAAAGATTAATCCGAGATCAGAAATG 240
 DB 304 CTTCGCTCTTTTACGAGTGGAGATAGAAAGATTAATCCGAGATCAGAAATG 363
 QY 241 AAGATTCAGTGGAGAGTTCCTCATGCTGAACACCTTGATCCCAAGACTGGAAGA 300
 DB 364 AAGATTCAGTGGAGAGTTCCTCATGCTGAACACCTTGATCCCAAGACTGGAAGA 423
 QY 301 ACCAAGATCATATATGACATCTCTGAGCTTGGCATGTGAGATACAAAGCTACAGAGAC 360
 DB 424 ACCAAGATCATATATGACATCTCTGAGCTTGGCATGTGAGATACAAAGCTACAGAGAC 483
 QY 361 AGATCAAGACAGCTCTAAGCAATGTTTAAACATCACCCAGCAACAGGAAAGAG 420
 DB 484 AGATCAAGACAGCTCTAAGCAATGTTTAAACATCACCCAGCAACAGGAAAGAG 543
 QY 421 CTGGGAGACCAATTAAGAGAGATTAATGACTACTGATCGATTAACCTTAAGTATG 480
 DB 544 CTGGGAGACCAATTAAGAGAGATTAATGACTACTGATCGATTAACCTTAAGTATG 603
 QY 481 AATGTATCTGATCAGTGTG 500
 DB 604 AATGTATCTGATCAGTGTG 623

RESULT 2
 ABR09776
 ID ABR09776 standard; CDNA; 1845 BP.
 XX

AC ABR09776;
 XX
 XX 14-MAR-2002 (first entry)
 DT
 XX Human ovarian tumour protein encoding cDNA #309.
 DE
 XX Human ovarian tumour protein; cancer; cytostatic; immunostimulant; ss;
 XX gene therapy; CD4+ T cell; CD8+ T cell; PCR primer.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200190154-A2.
 XX
 XX 29-NOV-2001.
 XX
 XX 23-MAY-2001; 2001WO-US16895.
 XX
 XX 24-MAY-2000; 2000US-207107P.
 XX 13-JUN-2000; 2000US-211457P.
 XX 21-JUN-2000; 2000US-213673P.
 XX 03-AUG-2000; 2000US-223288P.
 XX 01-MAR-2001; 2001US-272790P.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Xu J, Mitcham JL, Harlocker SL, Dillon DC, Secrist H, Lodes MJ;
 PI Algate PA, Fling SP, Mannion J, Benson DR, Carter D;
 XX
 XX WPI: 2002-097641/13.

PT New isolated polynucleotide encoding polypeptide comprising portion of
 PT ovarian tumour protein, useful for detection, diagnosis and therapy of
 PT human ovarian cancer -
 PT
 XX
 XX Claim 1: Page 256; 285pp; English.

XX The invention relates to an isolated polynucleotide encoding a
 CC polypeptide comprising a portion of an ovarian tumour protein. The
 CC sequences of the invention are useful for stimulating an immune response
 CC and for treating ovarian cancer in a patient. An antigen presenting cell
 CC that expresses the sequences is useful for treating ovarian cancer by
 CC incubating CD4+ and/or CD8+ T cells isolated from a patient. The T cells
 CC can then be proliferated and administered to the patient to inhibit the
 CC development of cancer. The DNA sequences are useful as probes or primers
 CC for nucleic acid hybridisation, to direct expression of a polypeptide in
 CC appropriate host cells. Detecting the presence of a cancer in a patient
 CC involves obtaining a biological sample from the patient, contacting the
 CC biological sample with an agent that binds to the protein, detecting the
 CC amount of protein that binds to the agent, comparing the amount of
 CC protein to a predetermined cut-off value and determining the presence of
 CC cancer. Sequences ABR09464-ABR09802 represent PCR primers and cDNA
 CC molecules encoding ovarian tumour proteins of the invention.
 CC
 XX Sequence 1845 BP; 723 A; 312 C; 413 G; 397 T; 0 other;

Query Match 100.0%; Score 500; DB 24; Length 1845;
 Best Local Similarity 100.0%; Pred. No. 1.9e-142;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGGTAGCGCGCGCCGACCGGGGTAGACCGGGGCGCATATATCTGCTGCCAA 60
 DB 1 GGGAGGTAGCGCGCGCCGACCGGGGTAGACCGGGGCGCATATATCTGCTGCCAA 60
 QY 61 GCGCGGAGCGCGCGGGGACCGGCATCACCCAGCTCTGACCTCTCTACACTCT 120
 DB 61 GCGCGGAGCGCGCGGGGACCGGCATCACCCAGCTCTGACCTCTCTACACTCT 120
 QY 121 GTCAAGTTGACCTGTGGAGATGTTTGAAGCTTTTGAAGAGGAGAAACGAATG 180
 DB 121 GTCAAGTTGACCTGTGGAGATGTTTGAAGCTTTTGAAGAGGAGAAACGAATG 180
 QY 181 CTTCGCTCTTTTACGAGTGGAGATTAAGAAAGATTAATCCGAGAAATCAGAAATG 240
 DB 181 CTTCGCTCTTTTACGAGTGGAGATTAAGAAAGATTAATCCGAGAAATCAGAAATG 240

Db 181 CTTCGCTCTTTTCAAGCACTGGAGATTAAGAAAGACTTATCCGAGAAATCAGAGATG 240
 QY 241 AAGAAATTCAGATTGGAGAGAGTTTCCATGCTGAAAAACATTGATCCCAAGACTGGAGAG 300
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC 241 AAGAAATTCAGATTGGAGAGAGTTTCCATGCTGAAAAACATTGATCCCAAGACTGGAGAG 300
 Db 301 ACCAAGATCATATATGACATGCTTGGACTTGGCCATGAGATATACAGGCTACACAGAGAC 360
 QY 301 ACCAAGATCATATATGACATGCTTGGACTTGGCCATGAGATATACAGGCTACACAGAGAC 360
 Db 301 AGATCAAAAGCAGCTCATATAAGCAATGGTTTAAACATCACCACAGACAAGGAAAGCAG 420
 QY 361 AGATCAAAAGCAGCTCATATAAGCAATGGTTTAAACATCACCACAGACAAGGAAAGCAG 420
 Db 361 AGATCAAAAGCAGCTCATATAAGCAATGGTTTAAACATCACCACAGACAAGGAAAGCAG 420
 QY 421 CTGGTGAACCAATTAAGAGAGATATGACTACTTCACTTGCATTAAGCTTATG 480
 Db 421 CTGGTGAACCAATTAAGAGAGATATGACTACTTCACTTGCATTAAGCTTATG 480
 QY 481 AATGTTATCTGATCCAGTG 500
 Db 481 AATGTTATCTGATCCAGTG 500

RESULT 3

AAD11115
 ID AAD11115 standard; DNA: 1860 BP.

AC AAD11115;

DT 24-SEP-2001 (first entry)

DE Human small cell lung cancer associated gene, MPP11.

KW Human; small cell lung cancer; therapy; hCAP; nucleic acid; NA;
 melanoma; cancer; colon; breast; head; neck; transitional cancer;
 leiomyosarcoma; helix-loop-helix; HLH; protein-binding factor; MPP11;
 synovial sarcoma; cytosolic; ds.

OS Homo sapiens.

PN WO200153349-A2.

PD 26-JUL-2001.

PE 19-JAN-2001; 2001WO-US02015.

PR 21-JAN-2000; 2000US-0489101.

PA (LUDWIG-) LUDWIG INST CANCER RES.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PI (CORR) CORNELL RES FOUND INC.

PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;

DR WPI; 2001-457597/49.

PT Isolated polypeptide, used to treat or prognose a disorder

PT characterised by expression of a hCAP e.g. cancer, is encoded by an

PS isolated nucleic acid comprising an NA Group 3 or 4 molecule -

PS Claim 57; Page 94-95; 152pp; English.

CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.

CC The present sequence is a DNA encoding human helix-loop-helix (HLH)
 CC protein-binding factor MPP11. This small cell lung cancer associated
 CC gene is designated as NY-SCLC-6.
 XX
 SX Sequence 1860 BP; 738 A; 312 C; 413 G; 397 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 1860;
 Best Local Similarity 100.0%; Pred. No. 1.9e-142;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGCTGAGCCGCTCCGCCACCGGGCTAGACCCGGCCATCATGCTTCTGCCAA 60
 Db 1 GGGAGCTGAGCCGCTCCGCCACCGGGCTAGACCCGGCCATCATGCTTCTGCCAA 60
 QY 61 GCGCGCGGAGCGCGCGGCGGACCCGATCACCAGCGCTGACCTGCGCTACACTCT 120
 Db 61 GCGCGCGGAGCGCGCGGCGGACCCGATCACCAGCGCTGACCTGCGCTACACTCT 120
 QY 121 GTCAGATTGAACCTGTGGGAAGATGTTTGAAGCTTTTGAAGAGAGAAACAGAAATG 180
 Db 121 GTCAGATTGAACCTGTGGGAAGATGTTTGAAGCTTTTGAAGAGAGAAACAGAAATG 180
 QY 181 CTTCGCTCTTTTCAAGCACTGGAGATTAAGAAAGCTTATCCGAGAAATCAGAGATG 240
 Db 181 CTTCGCTCTTTTCAAGCACTGGAGATTAAGAAAGCTTATCCGAGAAATCAGAGATG 240
 QY 241 AAGAAATTCAGATTGGAGAGAGTTTCCATGCTGAAAAACATTGATCCCAAGACTGGAGAG 300
 Db 241 AAGAAATTCAGATTGGAGAGAGTTTCCATGCTGAAAAACATTGATCCCAAGACTGGAGAG 300
 QY 301 ACCAAGATCATATATGACATGCTTGGACTTGGCCATGAGATATACAGGCTACACAGAGAC 360
 Db 301 ACCAAGATCATATATGACATGCTTGGACTTGGCCATGAGATATACAGGCTACACAGAGAC 360
 QY 361 AGATCAAAAGCAGCTCATATAAGCAATGGTTTAAACATCACCACAGACAAGGAAAGCAG 420
 Db 361 AGATCAAAAGCAGCTCATATAAGCAATGGTTTAAACATCACCACAGACAAGGAAAGCAG 420
 QY 421 CTGGTGAACCAATTAAGAGAGATATGACTACTTCACTTGCATTAAGCTTATG 480
 Db 421 CTGGTGAACCAATTAAGAGAGATATGACTACTTCACTTGCATTAAGCTTATG 480
 QY 481 AATGTTATCTGATCCAGTG 500
 Db 481 AATGTTATCTGATCCAGTG 500

RESULT 4

ABL16647
 ID ABL16647 standard; DNA: 2225 BP.

AC ABL16647;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 1414.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ds.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PA 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EM;
 XX WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1: SEQ ID NO 1414; 21bp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABH57737-ABH72072).
 CC (The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 2225 BP; 623 A; 560 C; 632 G; 410 T; 0 other;
 Query Match 16.0%; Score 80.2; DB 23; Length 2225;
 Best Local Similarity 59.4%; Pred. No. 3.3e-14;
 Matches 153; Conservative 0; Mismatches 103; Indels 3; Gaps 1;
 QY 224 CGAGGAATCAGAAATGAAAGATTTGCAAGTTTCCCATGCTGAAACACTTGA 283
 DB 306 CGATGAGAAATTTGAGGCGGTGGCGAGGAGGACATCATCTAAATGCTGGA 365
 QY 284 TCCCAAGATGGAAGACCAAGATCATTTGCAATTTGGCTTGGCATGAGATGA 343
 DB 366 TCTTAAGATGGAAGACCAAGATCATTTGCAATTTGGCTTGGCATGAGATGA 425
 QY 344 CAGGCTACACAGACAGATCAAGACGCTCAATAAGATGTTTAAACATCACC 403
 DB 426 CGAGGCGGAGGATGATTTGACGGCTTAAACGCGCATGTTCTGTCACCATCC 485
 QY 404 AGACAAACGGAAGACAGTGTGTAACCAATTAAGAGAGATTAATGACTTACTTCACTTG 463
 DB 486 CGATTAAGCGGAAAGCCAGGCGGAGGAGATCATCC--GGAGCATGATTACTTACATG 542
 QY 464 CATTAAGTAAAGCTTATGAAT 484
 DB 543 CATTAAGTAAAGCTTATGAAT 563
 RESULT 5
 AAH29904
 ID AAH29904 standard; DNA; 1876 BP.
 AC
 XX AAH29904;
 AC
 XX
 DT 27-JUL-2001 (first entry)
 XX
 DE C albicans apoptosis associated coding sequence #48.
 XX
 KM Yeast; fungus; apoptosis; infection; proliferative disease;
 KM vaccine; autoimmune disease; ischaemia; neurodegeneration; ds.
 XX
 OS *Candida albicans*.
 XX
 PN WO200102550-A2.
 XX
 PD 11-JAN-2001.
 XX
 PF 03-JUL-2000; 2000WO-BE00077.
 XX
 PR 01-JUL-1999; 99EP-0870141.
 XX
 PA (JANC) JANSSEN PHARM NV.

XX
 PI Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;
 XX WPI; 2001-367042/38.
 XX
 PT Yeast and fungal nucleic acids encoding proteins involved in a pathway
 PT leading to programmed cell death, useful for treating proliferative
 PT disorders, yeast and fungal infections, or for preventing apoptosis in
 PT certain diseases -
 XX
 PS Claim 17, Fig 2; 218pp; English.
 CC The present invention provides the protein and coding sequences of a
 CC number of apoptosis associated proteins from the yeast *Saccharomyces*
 CC cerevisiae and the fungus *Candida albicans*. These can be used to identify
 CC treatments for fungal and yeast infections, for proliferative diseases
 CC and for apoptosis related diseases such as autoimmune diseases, ischaemia
 CC and neurodegeneration. The present sequence is one of the *C. albicans*
 CC coding sequences of the invention.
 XX
 SO Sequence 1876 BP; 721 A; 294 C; 368 G; 493 T; 0 other;
 Query Match 12.1%; Score 60.4; DB 22; Length 1876;
 Best Local Similarity 60.2%; Pred. No. 5.6e-08;
 Matches 100; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
 QY 278 ACTTGATCCCAAGACTGGAAGAACCAAGATCATTTGCAATTTGGACTTGGCCATGT 337
 DB 743 ACAGATCCCAAGAGAAATGGAAGACTCCCAATTTATGCTGTTTAAAGTTTATCTCATTT 802
 QY 338 GAGATACAAAGCTTACACAGAGACATCAAGAGGCTCATTAAGCAATGTTTAAACA 397
 DB 803 GAGAGTAAAGCCATGACATCAATCAAGAGAGCCACAGAAAACAAGTTTGAACA 862
 QY 398 TCACCAGACAAAGGAAAGCAGCTGTGTAACCAATTAAGAGA 443
 DB 863 TCATCCAGATTAATAATCCGCTAGTGTGATTAAGAAACGATGA 908
 RESULT 6
 AAH29775
 ID AAH29775 standard; DNA; 1802 BP.
 AC
 XX AAH29775;
 AC
 XX
 DT 27-JUL-2001 (first entry)
 XX
 DE *S. cerevisiae* apoptosis associated coding sequence YGR284C.
 XX
 KM Yeast; fungus; apoptosis; infection; proliferative disease;
 KM vaccine; autoimmune disease; ischaemia; neurodegeneration; ds.
 XX
 OS *Saccharomyces cerevisiae*.
 XX
 PN WO200102550-A2.
 XX
 PD 11-JAN-2001.
 XX
 PF 03-JUL-2000; 2000WO-BE00077.
 XX
 PR 01-JUL-1999; 99EP-0870141.
 XX
 PA (JANC) JANSSEN PHARM NV.
 XX
 PI Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;
 XX WPI; 2001-367042/38.
 XX
 PR P-PSDB; AAG70739.
 XX
 PT Yeast and fungal nucleic acids encoding proteins involved in a pathway

PT leading to programmed cell death, useful for treating proliferative
PT disorders, yeast and fungal infections, or for preventing apoptosis in
PT certain diseases -

PS Claim 1; Fig 1; 218pp; English.

CC The present invention provides the protein and coding sequences of a
CC number of apoptosis associated proteins from the yeast *Saccharomyces*
CC *cerevisiae* and the fungus *Candida albicans*. These can be used to identify
CC treatments for fungal and yeast infections, for proliferative diseases
CC and for apoptosis related diseases such as autoimmune diseases, ischaemia
CC and neurodegeneration. The present sequence is one of the *S. cerevisiae*
CC coding sequences of the invention

Sequence 1802 BP; 593 A; 346 C; 368 G; 495 T; 0 other;

Query Match	10.4%	Score 51.8;	DB 22;	Length 1802;
Best Local Similarity	55.7%;	Pred. No. 2.3e-05;		
Matches 123; Conservative	0;	Mismatches 92;	Indels 6;	Gaps 1.

Qy	276	ACACATGATCCCAAGAAAGATGGAAACAAACAGATCATTAATGACAGTTGGACAT	353
Db	759	ACACATGATCTAGAGAGATGGAAACATGCCGATTTGTATGCTGATAGGTTTGCTAG	818
Qy	336	GTGAGATATCAAGCTACACAGAGACATCAAAAGCAAGCTCATAAAGCAATGTTTTAAA	395
Db	819	TTGCGTTTCAGAGCTACAGTCAAGTCAATCATCATCAAGGCTCACAGAAACAAAGTGTCAAG	878
Qy	396	CATCACCCAGCAAAAGCGAAGGAGCTGGGAAACCAATAAAGAAAGAGATATGACTAC	455
Db	879	TACACATCCAGCAAGCAATCTGCTGCTGGTGGTGAATTTGGACCAAG-----ATGGCTTT	932
Qy	456	TTTCACTTGCATATCAATAAGCTTATGAAAATGTTATGCAATCC	496
Db	933	TTTCAAGATTAATCAAAAGGCTTTGAAACCTTATGACGATTC	973

RESULT 7
AAT90969
ID AAT90969 standard; DNA; 3083 BP.

AC	AAT90969;
XX	
DT	22-JAN-1998 (first entry)
XX	

DE Zucotin gene.
XX Membrane-forming peptide; amphiphilic; monovalent metal cation; suture;
internal lining; slow-diffusion drug delivery vehicle; erythropoietin;
tissue-type plasminogen activator; haemoglobin; insulin; artificial skin
separatrix; matrix; dialysis membranes; filter; collagen; cell migration;
Alzheimer's disease; scrapie infection; therapy; proteolytic digestion;
acidic pH resistant; stomach acid; cell adhesion; cell monolayer; zucotin;
tissue culture; differentiated cell; stratified cell layer; ds.

OS *Saccharomyces cerevisiae*.
yx

FH	Key	Location/Qualifiers
FT	CDS	1292..2593
FT		/*tag= a
FT		/product= zuotin
vv		

PN	US670483-A.	
XX		
PD	23-SEP-1997.	
XX		
PE	28-DEC-1992;	92US-0973326.
XX		
PR	28-DEC-1992;	92US-0973326.
XX	30-NOV-1994;	94US-0346849.
XX		
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY	
XX		

PI Holmes T, Lockshin C, Rich A, Zhang S, et al.

DR WPI; 1997-479506/44.
DR P-PSDB; AAW30256.

DK P-PSDB; AAWS02
YY

PT Membranes formed by self-assembly of amphiphilic peptide(s) - useful
PT as bio:material(s), separation matrices, drug delivery vehicles,
PT etc.

PS Example 5; Column 29-34; 49pp; English
xx

This sequence represents the yeast znotin genes. Fragments of the encoded protein can be used as the amphiphilic peptides used in the macroscopic membrane (MM) of the invention. The MM is formed by self-assembly of the amphiphilic peptides (see AA030219-W03033) in an aqueous medium containing monovalent metal cations. The MM are stable in serum, are non-cytotoxic, and are useful in biomaterial applications, such as medical products (e.g. sutures), or internal linings. The MM are useful as slow-diffusion drug delivery vehicles for protein-type drugs, including erythropoietin, tissue-type plasminogen activator, synthetic haemoglobin and insulin. They can be used in applications, such as separation matrices (e.g. dialysis membranes). Collagen may be combined with the peptides to produce membranes for use as artificial skin. The MM may be used for making very thin, transparent fabric. Drugs which inhibit the self assembly of the peptides into filaments or filamentous membranes may be useful for treating Alzheimer's disease or scrapie infection. As they are resistant to proteolytic digestion and alkaline and acidic pH (such as stomach acid), drug delivery vehicles made of the MM could be taken orally. The charged residues and conformation of the MM promote cell adhesion and migration. The permeability of the MM also permits diffusion of small molecules, to the underside of cell monolayers, useful for tissue culture of differentiated cells and/or stratified cell layers.

Sequence 3083 BP; 1026 A; 578 C; 627 G; 852 T; 0 other;

Query Match	10.4%;	Score 51.8;	DB 18;	Length 3083;
Best Local Similarity	55.7%;	Pred. No. 3.1e-05;		
Matches 123; Conservative	0;	Mismatches 92;	Indels 6;	Gaps 1

QY	276	ACACCTGATCCCAAGAGACGTGGAGAACAAGATTCATTTTGGAGCTTTGGACTTGGCCAT	335
Db	1550	ACACATGATGCTAGAGACGTGAAAACAGCCGATTTGTATGCTGCTATAGGGTTTGTCTAAG	1609
QY	336	GATAGATTAACAGCTACACAGACAGATCAATTAAGACGCTCTAAAGCAATGGTTTAA	395
Db	1610	TTGCGTTTCAAGGCTACTGAAAGTCAATCATCAAGGCTCAGAAAACAAGTTTCAAG	1669
QY	396	CATCACCCAGACAAAAGGAAAGCGCTGGTGAACCAATTAAGAAAGAGATTAATCACTAC	455
Db	1670	TACCATTCACAAACACATTCGCTGCTGGTGGTGAATTGGACCAAG-----ATGGCTTT	1722
QY	456	TTTCACTTGCATTAACCTTAAGCTTATGAAGATTTATCTGATCC	496
Db	1724	TTTCAAGATTTTTCAAAAGGCTTTGAAACTTTGACTGATGTC	1764

RESULT 8	
AAZ11218	
ID	AAZ11218 standard; DNA; 3083 BP

AC	AAZ11218;
XX	
DT	09-NOV-1999 (first entry)

Human zuotin coding sequence.

KW Membran-forming peptide: cell clumping; macroscopic membrane;
KW amphiphilic peptide; slow-diffusion drug delivery system; cell growth;
KW artificial skin, separation matrix; artificial tissue; scrapie infection,
KW Alzheimer's disease; liver cirrhosis; kidney amyloidosis;
KW protein conformational disease; human; zoon; zoon; ss.


```

FT      from B. subtilis"
FT      39242..39904
FT      /tag= j
FT      /label= MG033
FT      /note= "Previously identified as MORF-20100, the
FT      encoded protein shows 35.90 percentage
FT      identity to glycerol uptake facilitator
FT      (glpF) from B. subtilis"
FT      CDS
FT      complement (39873..40514)
FT      /tag= k
FT      /label= MG034
FT      /note= "Previously identified as MORF-20101, the
FT      encoded protein shows 48.13 percentage
FT      identity to thymidylate kinase (tdk)
FT      from B. subtilis"
FT      CDS
FT      40543..41787
FT      /tag= l
FT      /label= MG035
FT      /note= "Previously identified as MORF-20102, the
FT      encoded protein shows 30.71 percentage
FT      identity to histidyl-tRNA synthetase (hiss)
FT      from Mycopacterium leprae"
FT      CDS
FT      complement (44751..46277)
FT      /tag= m
FT      /label= MG038
FT      /note= "Previously identified as MORF-20105, the
FT      encoded protein shows 46.83 percentage
FT      identity to glycerol kinase (glpK)
FT      from E. coli"
FT      CDS
FT      complement (46268..47422)
FT      /tag= n
FT      /label= MG039
FT      /note= "Previously identified as MORF-19831 and
FT      MORF-20106, the encoded protein shows 43.20
FT      percentage identity to glycerol-3-phosphate
FT      dehydrogenase (GOT2) from S. cerevisiae"
FT      CDS
FT      49377..49643
FT      /tag= o
FT      /label= MG041
FT      /note= "The encoded protein shows 48.86 percentage
FT      identity to phosphohistidinoprotein-hexose
FT      phosphotransferase (ptsH) from Mycoplasma
FT      capricolum"
FT      CDS
FT      50060..51520
FT      /tag= p
FT      /label= MG042
FT      /note= "Previously identified as MORF-19832 and
FT      MORF-20108, the encoded protein shows 41.92
FT      percentage identity to spermidine/
FT      putrescine transport ATP-binding protein
FT      (potA) from E. coli"
FT      CDS
FT      51525..52382
FT      /tag= q
FT      /label= MG043
FT      /note= "Previously identified as MORF-20110, the
FT      encoded protein shows 26.51 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein (potB) from E. coli"
FT      CDS
FT      52366..53220
FT      /tag= r
FT      /label= MG044
FT      /note= "Previously identified as MORF-20111, the
FT      encoded protein shows 29.45 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein C (potC) from E. coli"
FT      CDS
FT      54658..55605
FT      /tag= s
FT      /label= MG046
FT      /note= "Previously identified as MORF-20112, the
FT      encoded protein shows 36.60 percentage
FT      identity to sialoglycoprotease (gcp)
FT      from Pasteurella haemolytica"
FT      CDS
FT      complement (56970..58310)

```

```

FT      /tag= t
FT      /label= MG048
FT      /note= "Previously identified as MORF-19834,
FT      MORF-20114 and MORF-20115, the encoded protein
FT      shows 43.02 percentage identity to signal
FT      recognition particle protein (fth) from B.
FT      subtilis"
FT      CDS
FT      58117..59079
FT      /tag= u
FT      /label= MG049
FT      /note= "Previously identified as MORF-20114 and
FT      MORF-20115, the encoded protein shows 44.78
FT      percentage identity to purine-nucleoside
FT      phosphorylase (deod) from E. coli"
FT      CDS
FT      59083..59754
FT      /tag= v
FT      /label= MG050
FT      /note= "Previously identified as MORF-20117, the
FT      encoded protein shows 83.03 percentage
FT      identity to deoxyribose-phosphate aldolase
FT      (deoc) from Mycoplasma pneumoniae"
FT      CDS
FT      complement (64898..65731)
FT      /tag= w
FT      /label= MG056
FT      /note= "Previously identified as MORF-20122, the
FT      encoded protein shows 30.25 percent
FT      identity to the protein disclosed in
FT      GB:D26185_99 from B. subtilis"
FT      CDS
FT      complement (65713..66249)
FT      /tag= x
FT      /label= MG057
FT      /note= "Previously identified as MORF-20123, the
FT      encoded protein shows 36.90 percentage
FT      identity to the protein disclosed in
FT      GB:D26185_104 from B. subtilis"
FT      CDS
FT      81047..82597
FT      /tag= y
FT      /label= MG067
FT      /note= "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (SPase) from Staphylococcus aureus"
FT      CDS
FT      91065..91919
FT      /tag= z
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"

```

Query Match 8.5%; Score 42.4; DB 18; Length 580073;
 Best Local Similarity 54.5%; Pred. No. 0.36;
 Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

OY 343 ACAAGCTACACAGACAGATCAAGACGATCAAGCAATGTTTAAACATCACC 402
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 22434 AAAAGCGTAGTTCCAGACATTAAGAGCTTTGAAGAGTGAATGCAATATCACC 22493
OY 403 CAGACAAAGGAAAGCAAGCTGCTGTAACCAATTAAGAGAGATATGACTACTCATT 462
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 22494 CCGATCGTCATTAAGCAAAAATGAACTACTCAAAAACAAATGAGGAAAAGTTTAAAG 22553
OY 463 GCATACCTAAGCTTATGAATGTTATCTGATCAAG 498
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 22554 AGTTAATGAAGCATATGAACTCTAAGTATGAAG 22589

```

RESULT 12
 AAX61812
 ID AAX61812 standard; DNA: 1039 BP.
 XX
 AC AAX61812;

DT 19-JUL-1999 (first entry)

XX B. burgdorferi antigenic protein coding sequence, t5-14.nt.

XX Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

OS Borrelia burgdorferi.

XX WO9859071-A1.

XX 30-DEC-1998.

XX 18-JUN-1998; 98WO-US12718.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (MED-) MEDIMMUNE INC.

XX Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI; 1999-189980/16.

XX P-PSDB; AAY20115.

XX Claim 1: Page 202; 275pp; English.

XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.

XX Sequence 1039 BP; 526 A; 94 C; 212 G; 207 T; 0 other;

XX Query Match 8.2%; Score 41.2; DB 20; Length 1039;

XX Best Local Similarity 45.4%; Pred. No. 0.03; Mismatches 178; Indels 0; Gaps 0;

XX Matches 148; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

XX 124 AAGTGAACCTGTGGAGATGTTGAGCTTTGTTAGAGAGAGAAACAGAAATGCTT 183

XX 45 AAGGAAATTAAGAGATTTTATAGTTAGTTAGTCCAGCAAAAGTAAATTAATCTT 104

XX 184 CTGCTCTTTTCAGAACTGAGATTAAGAAAGTTATCCGAGAAATCAGAAATGAG 243

XX 105 CAAGTAGTTCAAAATGATGATTAATGCAAAATTAATGCAAGAGATGAGATTAATG 164

XX 244 AATTCAGTTGAGAGATTTCCATGCTGAAACATCTGATCCCAAAAGCTGGAAGACC 303

XX 165 AATTAATGAGGAGGATGATCTATTAACAGCAATGATGATTAATGATTAATGATCCCG 224

XX 304 AAGATCATTAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 363

XX 225 AAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 284

XX 364 TCAAAAGAGCTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 423

XX 285 AACCAAGAGCTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 344

XX 424 GTGAAACATTAATGAGATTAAT 449

XX 345 TAGAGAGAGAGAGAGAGAGAT 370

RESULT 13

AAK61811
ID AAK61811 standard; DNA; 1125 BP.

XX AAK61811;

AC 19-JUL-1999 (first entry)

XX B. burgdorferi antigenic protein coding sequence, t5-14.nt.

XX Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

OS Borrelia burgdorferi.

XX WO9859071-A1.

XX 30-DEC-1998.

XX 18-JUN-1998; 98WO-US12718.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (MED-) MEDIMMUNE INC.

XX Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI; 1999-189980/16.

XX P-PSDB; AAY20114.

XX Claim 1: Page 202; 275pp; English.

XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.

XX Sequence 1125 BP; 565 A; 101 C; 223 G; 236 T; 0 other;

XX Query Match 8.2%; Score 41.2; DB 20; Length 1125;

XX Best Local Similarity 45.4%; Pred. No. 0.032; Mismatches 178; Indels 0; Gaps 0;

XX Matches 148; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

XX 124 AAGTGAACCTGTGGAGATGTTGAGCTTTGTTAGAGAGAGAAACAGAAATGCTT 183

XX 128 AAGGAAATTAAGAGATTTTATAGTTAGTTAGTCCAGCAAAAGTAAATTAATCTT 187

XX 184 CTGCTCTTTTCAGAACTGAGATTAAGAAAGTTATCCGAGAAATCAGAAATGAG 243

XX 188 CAAGTAGTTCAAAATGATGATTAATGCAAAATTAATGCAAGAGATGAGATTAATG 247

XX 244 AATTCAGTTGAGAGATTTCCATGCTGAAACATCTGATCCCAAAAGCTGGAAGACC 303

XX 248 AATTAATGAGGAGGATGATCTATTAACAGCAATGATGATTAATGATTAATGATCCCG 307

XX 304 AAGATCATTAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 363

XX 308 AAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 367

XX 364 TCAAAAGAGCTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 423

XX 368 AACCAAGAGCTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 427

XX 424 GTGAAACATTAATGAGATTAAT 449

Case	US-08-455-073A-1	Sequence 1, Appl
27	66.2	13.2
28	4362	2

[illegible]

	Best Match	16.2%	Score 80.8%;	DB 4:	Length 4403765;
	Query Local Similarity	61.0%;	Pred. No. 3.7e-05;		
	Matches 183; Conservative	0;	Mismatches 112;	Indels 5;	Gaps 3.
OY	206	CCATTTCACAGGCGCGCGGGGCGCCCCCGGCACAGCGCCCGCGGGGCGCGAGGCCCTC	265		
Dh	3938160	CCATGTGCTCCCGCGCGCGCGGCTTGTGCGCGCATTTGCGCGCATTTGCGCGCGCACCGCCGCTG	39381010		
OY	266	CAGGTGACTTGCTCCCGGTCTCG--CGCGCGCCAGAAGTCCGCGCGGTCTACGG	323		
Dh	3938100	CGCGCGGTGCGCCCTCCCGCCCGAGATCTCCCGACCGCGGGGATGCGCGAGGCGCCGGTG	3938041		
OY	324	CGCGTCGCGCGCGTGTGTGCGCGCGTGTGTCGCGCGCCCGCGCCCTCTGCGCGCGCTCTA	383		
Dh	3938040	CGCGCGGTGTGCGCGCGGTGCTCCCGCGCGCCCGCGTCCCCTCGCGCGGTGTCGG	3937981		
OY	384	CGGTGAGC-AcAGAGCGCCCTGAAGAGGCTCGCGGGGCCCTCTCGCGCAACC CGCGCAGTGG	442		
Dh	3937980	CCATGCGCGCATTTGCGCGCGCGCACCGCGGTGCGCGCGCGCGCGCGCGCACTTGTGCGGCC	3937921		
OY	443	TGCGCGCGCG--CGCGCAGGCGCGGCCCCCGGCTCGCGCGCGCATATTCGCGCGGCGG	500		
Dh	3937920	TGCGCGCGCGTACCGCGCTTGTCCCGCGTTCGCGCATGCGCGCGCAAGTGTGCGCGCGCAAG	3937861		

RESULT 2

US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128.155
; EARLIER APPLICATION NUMBER: US 60/091.650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054.646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 15.1%; Score 75.6; DB 3; Length 152331;
Best Local Similarity 50.1%; Pred. No. 3.2e-05;
Matches 203; Conservative 0; Mismatches 195; Indels 7; Gaps 1;

DB 88 GGGGGTGGGGGGCTCAGTAAGTCTTCCCGCCGACGCTAGAGGCTAGCGCCGAGAACCCCT 147
DB 21846 GGGTGTGTGGGGTGGGNGSTNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 21905
QY 148 GCAAGTGGGGGCTGAGCTCACTCCGCTTCTTGGCTCCAGAGGCTGGCCGAGATCC 207
DB 21906 CCCCCCGCCCGCCGAGGCGGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 21965
QY 208 ATTCCAGAGCGGCGGCGGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCGCA 267
DB 21966 CCCCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 22025
QY 268 GGTGACTGTCTCCGAGTCTGCGCGCGCCGAGAGTCCGCGCGGCTGTGCGGCGCG 327
DB 22026 GCG-----CGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCGAC 22078
QY 328 TGGCGCGCTGTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAGGCT 387
DB 22079 CCCCCCGCCCGAGCC 22138
QY 388 GGAACAGAGCGGCTGAGAGCAGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
DB 22139 CG 22198
QY 448 GCGCGCGCGAGCG 492
DB 22199 CCG 22243

RESULT 3

US-09-050-863-2/c
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Payan, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Teet, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050.863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 14.9%; Score 74.4; DB 3; Length 2580;
Best Local Similarity 48.7%; Pred. No. 6.4e-05;
Matches 230; Conservative 0; Mismatches 241; Indels 1; Gaps 1;

QY 27 TCCCGGTGTTCTTCTGACAGCTGCTGCGCCCGCCCTTCCCGTGTGGCTGAGATCC 86
DB 1358 TCTCTGCTGCGCCCTGCTGCTGCTGCGCCCGCCCTGCTGCTGCTGCTGCTGCTGCTG 1299
QY 87 GGGGGGTGGGCGGCTCTAGAGTCTTCCCGCCGAGGCTCAGAGGCTCAGAGCCCG 146
DB 1298 TCTCTGCTGCGCCCTGCTGCGCCCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1239
QY 147 TCGAGTGGGGGCTGAGTCCAGTCCGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 206
DB 1238 TCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1179
QY 207 CATTCAGAGCG 266
DB 1178 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1119
QY 267 AGTGTGACTGTCTCCGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
DB 1118 TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1060
QY 327 TCGCGCGCGCTGCTGCG 386
DB 1059 CTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1000
QY 387 TGAACAGAGCGGCTGAGAGCAGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
DB 999 CTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 940
QY 447 CCGCGCGCGCGAGCG 498
DB 939 CTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888

RESULT 4

US-09-359-081-2/c
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,647
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fujita, Sharon M.
REGISTRATION NUMBER: 38,459
REFERENCE/DOCKET NUMBER: 1218.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2706
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-910-647--1

```

Query Match	14.9%	Score 74.4;	DB 4;	Length 9600;
Best Local Similarity	48.7%;	Pred. No. 5.9e-05;		
Matches 230; Conservative	0;	Mismatches 241;	Indels 1;	Gaps 1

[illegible]

RESULT 7
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921

GENERAL INFORMATION:
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/884,611
FILING DATE: 19920518
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear

US-07-884-811-15

Query Match	14.98;	Score 74.4;	DB 1;	Length 10596;
Best Local Similarity	48.78;	Pred. No. 5.9e-05;		
Matches 230; Conservative	0;	Mismatches 241;	Indels 1;	Gaps 1.

Oy	27	TCCGCGGTTTTCCTTGGACGCTGTGAGCCCCCCCCCTTCCCGGCTGGGGCTGGAGTACC	86
Db	2938	TCTCGCTCTGCCCCCTCTGCTGCTGCCCCCTCTCTGCTCTGCCCCCTCTGCCCCCTTC	2879
Oy	87	GGGGGGTGGGGGCTCATGAACTCTTCCCGCACCTTAGGGTACGCGCCAGAAACCCC	146
Db	2878	TGCTCTGCCCCCTCTGCCCCCTCTGCTGCTGCCCCCTCTGCCCCCTCTGCTCTGCCCC	2819
Oy	147	TGCAGGTGCGGGGCGAGCTCTCAGTCCCGCTTCTTTCCTCTCCAGAGGATCGCCAGAGTTC	206
Db	2818	TCTGCCCCCTCTGCTCTGCTCTGCCCCCTCTGCCCCCTCTCTCGTCTCTGCCCCCTCTGCTC	275
Oy	207	CATTCCAGCGCGCGCGCGCCCCCGCCACAGTCCCAAGTCCCGGCGGCCAGCCCTTTC	266
Db	2758	CCCTCTGCTCTGCCCCCTCTGCTGCCCCCTCTGCTCTGCCCCCTCTGCCCCCTCTGCTCTC	2699
Oy	267	AGGTGAGATTGCTACCTCCGAGTCTGCGCGCGCCAGAGATCCCGCGGCTGCTGCGGCTCG	326
Db	2698	TGCCCCCTCTGCTCTGCCCCCTC-GTGTCTGCCCCCTCTGCTGCTGCCCCCTCTGCCCC	26410
Oy	327	CTGCGCGCGCTGCGCGCGCGGTGTGCGCGCGCCCCCGGCCCTTGCGCGCGCTCTACGG	386
Db	2639	CTCTGCGCCCCCTCTGCTCTGCGCCCCCTCTGCTCTGCTCTGCCCCCTCTGCCCCCTCTGCCC	2580
Oy	387	TGACACAGAGGCGCTGTAAACACACTCTCGCGGCGCCCCCTCGCGACCCCGGCGAGTGTGG	446
Db	2579	CTCTGCTCTGCGCCCCCTCTGCTCTGCTCTGCCCCCTCTGCCCCCTCTGCCCCCTCTGCTG	2520
Oy	447	CGCCCCGCGCGAGCGCGCGCCCCCGGCTCTGCGCGCCACTATGCGCGGCGGC	498

	Query Match	14.9%	Score 74.4	DB 1	Length 10596
	Best Local Similarity	48.7%	Pred. No. 5.9e-05		
	Matches 230;	Conservative	0;	Mismatches 241;	Indels 1; Gaps 1
QY	27	TCCCGGTGTTTCCCTCTGACADGCTGTCGTCGCCCCCCCCCTTCCCGCTGCTGGGACCTCGAGATCCC	86		
Db	2938	TCTCTCTCTGACCCCTCCCTGCTCTCTGCCCCCTTCTCTCTCTGCTCTGCCCCCTCTGCCCCCTTC	2879		
QY	87	GGGGGTGGGGCGGCGCTCATGAACCTCTTCCGCGCACTCAGGGTCAACGCCAGAACCCCC	146		

Query Match	14.9%	Score 74.4;	DB 2;	Length 10596;
Best Local Similarity	48.7%;	Pred. No. 5.9e-05;		
Matches 230; Conservative	0;	Mismatches 241;	Indels 1;	Gaps 1


```

QY 27 TCCCGGTGTTCTTGCAGACGTCGTCGCCCCCCCCCTTCCCGCTGCTGGGCTGACATCCC 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2938 TCCGTCTGCTGCCCCCTCTGCTGCTGCCCCCTCTCTGCTGCTGCCCCCTCTGCTGCTG 2879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 GGGGGGTGGGGGCTCATGAACTCTTCCCGCACCTCAGGGTCACGCCAGAACCCCC 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2878 TGTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 147 TGCAGGTGCGGGGCTGAGCTCAGTCCCGCTTGTGGCTCCAGGGCTGCGCCAGAGTTC 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2818 TCTGCCCCCTCTGCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCC 2759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 207 CATTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2758 CCGCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 AGGTGACTTGTCTCCGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2698 TGCCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 327 CTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2639 CTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 TGCACACAGCGCGCTGAAGACCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2579 CTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2519 CTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12

PCT-US93-04648-15/c
Sequence 15, Application PC/TUS9304648

GENERAL INFORMATION:

APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melanie H.
TITLE OF INVENTION: HEPATOCTYPE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patlin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648
FILING DATE: 19930517

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885971
FILING DATE: 18-MAY-92

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755,779P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US93-04648-15

Query Match 14.9%; Score 74.4; DB 5; Length 10596;
Best Local Similarity 48.7%; Pred. No. 5,9e-05;
Matches 230; Conservative 0; Mismatches 241; Indels 1; Gaps 1;

```

QY 27 TCCCGGTGTTCTTGCAGACGTCGTCGCCCCCCCCCTTCCCGCTGCTGGGCTGACATCCC 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2938 TCCGTCTGCTGCCCCCTCTGCTGCTGCCCCCTCTCTGCTGCTGCCCCCTCTGCTGCTG 2879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 GGGGGGTGGGGGCTCATGAACTCTTCCCGCACCTCAGGGTCACGCCAGAACCCCC 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2878 TGTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 147 TGCAGGTGCGGGGCTGAGCTCAGTCCCGCTTGTGGCTCCAGGGCTGCGCCAGAGTTC 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2818 TCTGCCCCCTCTGCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCC 2759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 207 CATTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2758 CCGCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 AGGTGACTTGTCTCCGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2698 TGCCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 327 CTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2639 CTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 TGCACACAGCGCGCTGAAGACCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2579 CTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2519 CTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTGCCCCCTCTGCTGCTG 2468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13

US-09-056-556-182/c
Sequence 182, Application US/09056556

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-Apr-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Mark, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457

TRE

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 985 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-182

Query Match
Best Local Similarity 51.0%; Score 73.2; DB 4; Length 985;
Matches 198; Conservative 0; Mismatches 188; Indels 2; Gaps 1;

115 CCGCCACCTGACGAGTGCAGACCCGAGACCCGAGTGGGCTGAGCTCAGTCCG 174
116 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
117 CCGCCACGAGCAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
118 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
119 CTTCTTGTCTCCAGGCGCTGGCGCCAGATTCATTCAGCGCGCGCGCGCGCGCG 234
120 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 CTTGGCTGTAGGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
122 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
123 CACGCCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 292
124 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
125 GTCACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454
126 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
127 GCGCGCCAGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 352
128 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
129 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
130 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
131 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 412
132 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
133 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 334
134 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
135 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 472
136 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
137 GCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274
138 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
139 GCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
140 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
141 GGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
142 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-165-264-11/c
Sequence 11, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinayagamorthy, Thuralayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44/47
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 11
LENGTH: 320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

Query Match
Best Local Similarity 51.7%; Score 72.6; DB 4; Length 320;
Matches 165; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

177 TCTTGGCTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 236
178 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
179 TCTTGTGTAGTGCAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
180 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
182 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
183 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
184 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
185 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381
186 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
187 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
188 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
189 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
190 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
191 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
192 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
193 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
194 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
195 TGTCTTGTTCGAAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
196 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
197 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
198 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

260 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 201
261 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
262 CCGAGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
263 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
264 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 141
265 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
266 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
267 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
268 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
269 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
270 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
271 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
272 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21
273 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
274 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
275 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
276 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21
277 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
278 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
279 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
280 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21
281 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-09-165-264-14/c
Sequence 14, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinayagamorthy, Thuralayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44/47
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

Query Match
Best Local Similarity 51.7%; Score 72.6; DB 4; Length 320;
Matches 165; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

177 TCTTGGCTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 236
178 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
179 TCTTGTGTAGTGCAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
180 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
182 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
183 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
184 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
185 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381
186 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
187 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
188 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
189 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
190 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
191 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
192 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
193 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
194 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
195 TGTCTTGTTCGAAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
196 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
197 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
198 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: October 10, 2002, 18:13:55
Job time : 1173.73 secs

vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com>

BASE COUNT 126 a 371 c 266 g 182 t 9 others
ORIGIN

Query Match 63.3% Score 316.6 DB 9: Length 954;
Best Local Similarity 94.2% Pred. No. 1.8e-30;
Matches 325; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

QY 2 GGCCTAGCGGGGGGCGAGCCATGTCCTCCGCTGTCCTGACGCTGAGCCGCC 61
DB 103 GGGCCGCTGGCGGAGGCGCATGTCCTCCGCTGTCCTGACGCTGAGCCGCC 162
QY 62 TTCCCGCTGCTGCGGCTGAGCTCCGCGGGGGGGGCTGATGACCTCTCCGCCA 121
DB 163 TTCCCGCTGCTGCGGCTGAGCTCCGCGGGGGGGGCTGATGACCTCTCCGCCA 222
QY 122 CCTCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 181
DB 223 CCTCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 282
QY 182 GCTTCCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 241
DB 283 GCTTCCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 342
QY 242 CAGGCGCGGGGCGGCGGAGCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 301
DB 343 CAGGCGCGGGGCGGCGGAGCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 402
QY 302 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 346
DB 403 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 447

RESULT 4

AL561632

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 921

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DL008YEO9"

/clone_1b="LTI_NFL010_BC2"

/sex="male"

/tissue_type="B cells from Burkitt lymphoma"

/note="Vector: pCMVSPORT 6; Site: 1; NotI: 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com>
BASE COUNT 105 a 375 c 286 g 141 t 14 others
ORIGIN

Query Match 61.8% Score 309.2; DB 9: Length 921;
Best Local Similarity 88.0% Pred. No. 1.5e-29;
Matches 331; Conservative 5; Mismatches 40; Indels 0; Gaps 0;

QY 2 GGCCTAGCGGGGGGCGAGCCATGTCCTCCGCTGTCCTGACGCTGAGCCGCC 61
DB 276 GGGCCGCTGGCGGAGGCGCATGTCCTCCGCTGTCCTGACGCTGAGCCGCC 335
QY 62 TTCCCGCTGCTGCGGCTGAGCTCCGCGGGGGGGGCTGATGACCTCTCCGCCA 121
DB 336 TTCCCGCTGCTGCGGCTGAGCTCCGCGGGGGGGGCTGATGACCTCTCCGCCA 395
QY 122 CCTCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 181
DB 396 CCTCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 455
QY 182 GCTTCCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 241
DB 456 GCTTCCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 301
QY 242 CAGGCGCGGGGCGGCGGAGCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 342
DB 516 CAGGCGCGGGGCGGCGGAGCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 402
QY 302 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 346
DB 576 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 402
QY 362 CCGGCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 447
DB 636 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 447

RESULT 5

BE514418

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 902

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3634139"

/clone_1b="NIH_MGC_8"

/tissue_type="Burkitt lymphoma"

/note="Vector: pCMVSPORT 6; Site: 1; NotI: 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com>
BASE COUNT 105 a 375 c 286 g 141 t 14 others
ORIGIN

Query Match 61.8% Score 309.2; DB 9: Length 921;
Best Local Similarity 88.0% Pred. No. 1.5e-29;
Matches 331; Conservative 5; Mismatches 40; Indels 0; Gaps 0;

QY 2 GGCCTAGCGGGGGGCGAGCCATGTCCTCCGCTGTCCTGACGCTGAGCCGCC 61
DB 276 GGGCCGCTGGCGGAGGCGCATGTCCTCCGCTGTCCTGACGCTGAGCCGCC 335
QY 62 TTCCCGCTGCTGCGGCTGAGCTCCGCGGGGGGGGCTGATGACCTCTCCGCCA 121
DB 336 TTCCCGCTGCTGCGGCTGAGCTCCGCGGGGGGGGCTGATGACCTCTCCGCCA 395
QY 122 CCTCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 181
DB 396 CCTCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 455
QY 182 GCTTCCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 241
DB 456 GCTTCCAGGCTCAGCGCCAGAACCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 301
QY 242 CAGGCGCGGGGCGGCGGAGCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 342
DB 516 CAGGCGCGGGGCGGCGGAGCCCTGCAAGTGGGGCTGAGCTCCAGTCCGCTTTT 402
QY 302 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 346
DB 576 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 402
QY 362 CCGGCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 447
DB 636 GAGTCCGCGGGGCTGCTGCGCGCGCTGCGCGCTGCTGCGCGC 447

Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalobosbcm.tmc.edu
Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Munz, D.M., Gibbs, R.A.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/ILND at: <http://image.llnl.gov>
Series: IRAC Plate: 6 Row: K Column: 1
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1. 2152
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3154686"
/tissue_type="Mammary tumor. Brcal-/fl: MMTV-Cre model. 10 months old, gross tissue."
/clone_1lb="NCL_CGAP_Mam3"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 435 a 692 c 619 g 406 t
ORIGIN

Query Match 51.48; Score 256.8; DB 11; Length 2152;
Best Local Similarity 87.08; Pred. No. 3.5e-23;
Matches 282; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 125 CAGGGTCACGCCACGACCCCTCAGAGTCGGGGCTGAGTCAGTCCCGCTTTTCC 184
DB 67 CAGGGTCACGCCACGACCCCTCAGAGTCGGGGCTGAGTCAGTCCCGCTTTTCC 126
QY 185 TCCACAGGGCTGGCCGACAGTTCATTCAGAGCCGCGCGCCGCCACGCCACG 244
DB 127 TCCACAGGGCTGGCCGACAGTTCATTCAGAGCCGCGCGCCGCCACGCCACG 186
QY 245 GCCCGCGCGCGCGAGCCCTCAGAGTGAAGTCTCCCGCTGCTCCGCCGCCGCCACG 304
DB 187 GCCCGCGCGCGCGAGCCCTCAGAGTGAAGTCTCCCGCTGCTCCGCCGCCGCCACG 246
QY 305 TCCCGCGCGCGCGAGCCCTCAGAGTGAAGTCTCCCGCTGCTCCGCCGCCGCCACG 364
DB 247 TCAGCGCGCGCGCGCGCGCGCTGCTGCGCGCGCTGCTCCAGTGTACTGACACCCG 306
QY 365 GCCCTGCGCGCGCGCTCAGAGTGAAGTCTCCCGCTGCTCCGCCGCCGCCACG 424
DB 307 GCCCTGCGCGCGCGCTCAGAGTGAAGTCTCCCGCTGCTCCAGTGTACTGACACCCG 366
QY 425 CCGCAGCCGCCCGCAGTGTGCGCG 448
DB 367 CAGCGCGCGGTGGCGCGCGCGCGCG 390

RESULT 8

BE988668 379 bp mRNA linear EST 05-OCT-2000
LOCUS BE988668/c
DEFINITION UI-M-CGDP-bhy-g-11-0-UI-s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
ACCESSION BE988668
VERSION BE988668.1 GI:10665238
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Bonaldi, M.F., Lennon, G., and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H

FEATURES

Location/Qualifiers
1. 379
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CGDP-bhy-g-11-0-UI"
/clone_1lb="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_Ret4_S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_Seq=None found"
BASE COUNT 46 a 109 c 181 g 43 t
ORIGIN

Query Match 49.68; Score 248.2; DB 10; Length 379;
Best Local Similarity 87.78; Pred. No. 5e-22;
Matches 271; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 174 GCTTTCTTCCCTCCAGAGGCTGCGCCAGAGTCCATTCAGAGCCGCGCGCGCCCGC 233
DB 309 GCAGCTTATCTTTCAGAGGCTGCGCCAGAGTCCATTCAGAGCCGCGCGCGCCCGC 250
QY 234 CCAGCGCCCGAGGCG 293
DB 249 CCAGCGCCCGAGGCG 190
QY 294 CCGCCAGAGTCCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353
DB 189 CCGCGAGGAATCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130
QY 354 CCG 413
DB 129 CTGCAACCCCG 70
QY 414 CG 473
DB 69 CGCGCGCTCTCCG 10
QY 474 CCTCGCGCG 482
DB 9 CCGCGCGCG 1

RESULT 9
BG773423 465 bp mRNA linear EST 15-MAY-2001
LOCUS BG773423
DEFINITION 602720325F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4837135 5',
ACCESSION BG773423
VERSION BG773423.1 GI:14084076

Result No.	Score	Query Match	length	DB	ID	Description
1	500	100.0	2389	22	AAD11114	Human small cell 1
2	90.6	18.1	2561	22	AAH26500	Rabbit low density
3	81.2	16.2	37856	21	AAAI1992	S. cellulosum DNA
4	80.8	16.2	4403765	22	AAI199683	Mycobacterium tube
5	78.8	15.8	1614	22	AAH26499	Human low density
6	78.8	15.8	12425	22	AAH26495	Human low density
7	77.4	15.5	114955	20	AAH53491	Human adenosine AI
8	77.2	15.4	1337	20	AAH217261	Human gene express
9	75.6	15.1	3198	20	AAH02974	Human IL-1ra BAC c

C 10	74.4	14.9	799	19	AAV55831	Nucleotide sequenc
C 11	74.4	14.9	1926	21	AAAS0254	Epsstein Barr virus
C 12	74.4	14.9	1926	22	AAAF82902	EBV tethering prot
C 13	74.4	14.9	2580	21	AAAF75454	Nucleotide sequenc
C 14	74.4	14.9	5452	20	AAAG00923	Anti-sense strand
C 15	74.4	14.9	8705	20	AAZ23776	Vector pShuttle DN
C 16	74.4	14.9	9600	19	AAVZ1683	Vector plasmid pCM
C 17	74.4	14.9	10380	20	AAZ22248	Nucleotide sequenc
C 18	74.4	14.9	10596	14	AAOS1731	Plasmid pCisEBON f
C 19	74.4	14.9	10596	17	AAAT40348	Plasmid pCisEBON f
C 20	74.4	14.9	10596	20	AAAT15650	Nucleotide sequenc
C 21	74.4	14.9	16080	21	AAAS9553	DNA clone pCEK Cl.
C 22	73.2	14.6	985	19	AAAF64548	M. tuberculosis im
C 23	73.2	14.6	985	19	AAV44439	Mycobacterium tube
C 24	73.2	14.6	985	20	AAZ19349	M. tuberculosis an
C 25	73.2	14.6	985	20	AAZ19137	M. tuberculosis re
C 26	72.6	14.5	320	21	AAAS3183	Primer used in the
C 27	72.6	14.5	320	21	AAAS38186	Primer used in the
C 28	72.6	14.5	1000	21	AAAS02484	Human colon cancer
C 29	72.6	14.5	114955	20	AAAS3491	Human adenosine A1
C 30	71.2	14.2	320	21	AAAS38185	Primer used in the
C 31	70.8	14.2	2668	22	AAK75409	Human immune/haema
C 32	70.6	14.1	318	21	AAAS38184	Primer used in the
C 33	70.2	14.0	1028	13	AAQ27091	XY26 probe. Homo
C 34	70.2	14.0	1028	22	AAAS15942	1.0kb pSLT1 fragmen
C 35	70.2	14.0	1028	22	AAAS01347	Human Frangible X Sy
C 36	69.8	14.0	1102	22	AAAS16923	Human gene express
C 37	69.2	13.8	4411529	12	AAI196682	Mycobacterium tube
C 38	69	13.8	8438	15	AAO73500	DNA encoding Pseud
C 39	68.8	13.8	955	22	AAAS01024	Sugarcane plant ge
C 40	68.8	13.8	1007	22	AAAS01022	Sugarcane plant ge
C 41	68.8	13.8	1698	23	AAAS84408	DNA encoding novel
C 42	68.8	13.8	2049	23	AAAF6210	DNA encoding novel
C 43	68.6	13.7	12001	16	AAO76213	HSV L/STF region.
C 44	68	13.6	795	19	AAO55830	FGA Insert Stabli
C 45	68	13.6	2188	20	AAZ77506	Human ovarian tumo

ALIGNMENTS

XX	RESULT 1
XX	AAD11114
ID	AAD11114 standard; DNA; 2389 BP.
XX	
AC	AAD11114;
XX	
DT	24-SEP-2001 (first entry)
DE	
XX	Human small cell lung cancer associated gene, MAZ.
KW	Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA
KW	melanoma; cancer; colon; breast; head; neck; transitional cancer
KW	leiomyosarcoma; myc-associated Zinc-finger protein; MAZ; cytosta
KW	synovial sarcoma; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200153349-A2.
XX	
PD	26-JUL-2001.
XX	
PF	19-JAN-2001; 2001WO-US02015.
XX	
PR	21-JAN-2000; 2000US-0489101.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
DR	WPI: 2001-457597/49.

xx Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e.g. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 xx
 PS Claim 57; Page 92-94; 152pp; English.

xx The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterised by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterised by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterised by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is a DNA encoding human myc-associated Zinc-finger
 CC protein MZ. This small cell lung cancer associated gene is designated
 CC as NY-SCLC-5.
 xx

xx Sequence 2389 BP; 419 A; 805 C; 731 G; 434 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 1.5e-63;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCTACGGGGGGGCGAGCCATTTCCCGGTGTTTCTTTCACAGCTGTCGCCCCC 60
 DB 1 CGGCTACGGGGGGGCGAGCCATTTCCCGGTGTTTCTTTCACAGCTGTCGCCCCC 60
 QY 61 CTTCCCGGCTGGGCTGAGACTCCGGGGGGTGGGGCCCATGAACTCTTCCCGCC 120
 DB 61 CTTCCCGGCTGGGCTGAGACTCCGGGGGGTGGGGCCCATGAACTCTTCCCGCC 120
 QY 121 ACCTCAGGCTCAGCCCAAGACCCCTTGCAGGTGCGGGCTGAGCTCCAGTCCCGCTTCT 180
 DB 121 ACCTCAGGCTCAGCCCAAGACCCCTTGCAGGTGCGGGCTGAGCTCCAGTCCCGCTTCT 180
 QY 181 TGCCCTCCAGGGCTGCGCCAGAGTCATTCAGGCGCGGGGGCCCGCCGCAAGCC 240
 DB 181 TGCCCTCCAGGGCTGCGCCAGAGTCATTCAGGCGCGGGGGCCCGCCGCAAGCC 240
 QY 241 CCAGGCCCCGGGGCGGAGCCCTTCAGGTGACTTGTCCCGGTGCTCGCCGCCCA 300
 DB 241 CCAGGCCCCGGGGCGGAGCCCTTCAGGTGACTTGTCCCGGTGCTCGCCGCCCA 300
 QY 301 GGAGTCCGGCGGCTGCTGCGGGCGGCTGCGCCGCTGCTGCGCCGCTGCGCGCC 360
 DB 301 GGAGTCCGGCGGCTGCTGCGGGCGGCTGCGCCGCTGCTGCGCCGCTGCGCGCC 360
 QY 361 CCGGCGCCCTGCGCGCTTACAGGTGACACACAGCGGCCCTGAAGACAGCTTCGCGCC 420
 DB 361 CCGGCGCCCTGCGCGCTTACAGGTGACACACAGCGGCCCTGAAGACAGCTTCGCGCC 420
 QY 421 CCGTCCGCGACCCCGGAGTGTGGCGCCCGGCGCCAGAGCCGCGCCCGCTCGCC 480
 DB 421 CCGTCCGCGACCCCGGAGTGTGGCGCCCGGCGCCAGAGCCGCGCCCGCTCGCC 480
 QY 481 CGCCACTATGCGCGCGCG 500
 DB 481 CGCCACTATGCGCGCGCG 500

RESULT 2

AAH26500 standard; cDNA; 2561 BP.

AAH26500:

12-NOV-2001 (first entry)

XX

DE Rabbit low density lipoprotein binding protein 2 (LBP-2) cDNA.

XX Low density lipoprotein binding protein 2; LBP-2; LDL; rabbit;
 KM atherosclerosis; antiatherosclerotic; gene therapy; diagnosis;
 KM vaccine; ss.

OS Oryctolagus cuniculus.

XX Key Location/Qualifiers

FT CDS 246..1928

FT /tag= a

PN W0200164874-A2.

XX 07-SEP-2001.

XX 28-FEB-2001; 2001WO-US06356.

XX 02-MAR-2000; 2000US-0517849.

XX 14-JUL-2000; 2000US-0616289.

XX (BOST-) BOSTON HEART FOUND INC.

XX Lees AM, Lees RS, Law SW, Arjona AA;

XX MPT: 2001-565505/63.

XX P-PSDB; AAB82807.

XX New isolated low density lipoprotein binding polypeptide for treating,

XX diagnosing and/or identifying therapeutic agents for atherosclerosis,

XX Claim 4; Fig 2A; 143pp; English.

XX The present sequence is that of cDNA encoding novel rabbit

XX low density lipoprotein binding protein 1 (LBP-2, see AAB82807).

XX The cDNA was isolated following screening of a rabbit cDNA library

XX for clones encoding LBPs that bound to both native low density

XX lipoprotein (LDL) and methyl LDL. The invention provides claimed

XX polynucleotides encoding novel polypeptides which are capable of

XX binding to native and methylated LDL, the isolated polypeptides,

XX termed LBPs, and biologically active fragments and analogues of

XX them, as well as expression vectors, cells and methods of producing

XX the LBPs. Also claimed are methods of determining if an animal is

XX at risk for atherosclerosis, methods for evaluating an agent for

XX use in treating atherosclerosis, and methods for treating a cell

XX having an abnormality in structure or metabolism of LBP.

XX Pharmaceutical compositions comprising an LBP polypeptide or

XX nucleic acid, and vaccine compositions, are also claimed.

XX Sequence 2561 BP; 372 A; 937 G; 879 C; 373 T; 0 other;

Query Match 18.1%; Score 90.6; DB 22; Length 2561;
 Best Local Similarity 54.8%; Pred. No. 2.7e-05;
 Matches 200; Conservative 0; Mismatches 164; Indels 1; Gaps 1;

QY 133 CGCCAGAACCCCTGCAAGTGGGGCTGAGTCCAGTCCGCTCTTTCCTCCAGAG 192
 DB 614 CGCCAGAACCCCTGCAAGTGGGGCTGAGTCCAGTCCGCTCTTTCCTCCAGAG 192
 QY 193 CTGCGCCAGAGTCCATTCCAGAGCCGCGCGCCGCGCCGCGCCAGAGCCCGCGGC 252
 DB 674 CGCCCGC 733
 QY 253 GCGCGAGCCCTCCAGGTGAGTCTTCCCGGTGCTGCGCGCGCGCGCGAGTCCGCGGC 312
 DB 734 GC 793
 QY 313 GCGTCTGCGCGCGCTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGC 372
 DB 794 GC 853
 QY 373 CGCCGCTCTAGGTGAGACAGCGCCCTGAAGAGGCTCTCGGCGCGCGCGCGCGC 432

```

Db 854 GCCCGCCGCCGCTCGCCGCCGAGTGCCTCGCCGCCGCCGACAGCCGCCGCC 913
QY 433 CCCGCAAGTGTGGCGCCGCCGAGGCCGCCGCCGCTCGCCGCACTATGCG 492
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 914 GCCCGCAAG-CAGCAGCAGAGCGCCGCCGCCGACGCGCCGACAGCAGCGCG 972
QY 493 GCGCG 497
    ||| |
Db 973 GCGCG 977

RESULT 3
AA11992/c
ID AA11992 standard; DNA: 37856 BP.
XX
XX AA11992:
XX
XX 07-AUG-2000 (first entry)
XX
XX S. cellulosum DNA encoding polyketide and heteropolyketide enzymes.
XX
XX Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
XX epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
XX plant-protection; ds.
XX
XX Sorangium cellulosum.
XX
XX Key location/Qualifiers
XX CDS complement (3398..6100)
XX FT /*tag- a
XX FT /product= "ORF1-1RNA synthetase"
XX FT /note= "gtg start codon"
XX FT complement (6374..7111)
XX FT /*tag- b
XX FT /product= "ORF2-monoxygenase"
XX FT complement (8433..9550)
XX FT /*tag- c
XX FT /product= "ORF3-aminotransferase"
XX FT /note= "AGT start codon given in the specification"
XX FT 9855..11393
XX FT /*tag- d
XX FT /product= "ORF4- tyrosine/DOPA-Decarboxylase"
XX FT /note= "GTG start codon"
XX FT 12212..13658
XX FT /*tag- e
XX FT /product= "ORF5-3-oxoacyl-ACP-reductase"
XX FT /note= "ACC start codon"
XX FT 15374..13984
XX FT /*tag- f
XX FT /product= "ORF6-polyketide synthase"
XX FT 20003..27889
XX FT /*tag- g
XX FT /product= "ORF7-peptide synthetase"
XX FT 28251..29400
XX FT /*tag- h
XX FT /product= "ORF8-transpeptidase"
XX FT complement (30040..31720)
XX FT /*tag- i
XX FT /product= "ORF9-regulation element"
XX FT /note= "CGC stop codon"
XX FT 31982..32932
XX FT /*tag- j
XX FT /product= "ORF10-transcription regulator"
XX FT 33128..33613
XX FT /*tag- k
XX FT /product= "ORF11-regulation element"
XX FT /note= "GTG start codon"
XX FT 33661..34077
XX FT /*tag- l
XX FT /product= "ORF12-regulation element"
XX FT complement (35255..35616)
XX FT /*tag- m
XX FT /product= "ORF13-transcription regulator"

```

```

FT CDS complement (35730..36242)
FT /*tag- n
FT /product= "ORF14-transcription regulator"
FT /note= "GTG start codon"
PN DE19846493-A1.
XX
PD 13-APR-2000.
XX
XX 09-OCT-1998; 98DE-1046493.
XX
XX 09-OCT-1998; 98DE-1046493.
XX
XX (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
XX Beyer S, Mueller R;
XX
XX WPI, 2000-294101/26.
XX
XX DNA sequence coding for products involved in the biosynthesis of
XX polyketide or heteropolyketide compounds, especially epothilone
XX
XX Claim 3: Page 20-33; 36pp; German.
XX
XX This invention describes a novel DNA sequence (I) whose expression
XX products effect or are involved in the enzymatic biosynthesis,
XX mutasynthesis or partial synthesis of polyketide or heteropolyketide
XX compounds (II). (I) can be inserted into an expression vector and used
XX to transform or transfect prokaryotic or eukaryotic cells with the aim
XX of obtaining strains that produce large amounts of polyketide or
XX heteropolyketide compounds, especially epothilones, which have cytotoxic
XX and/or immunosuppressant and antibiotic and antifungal activities and
XX are useful as plant-protection agents. This sequence represents the DNA
XX sequence isolated from Sorangium cellulosum which is described in the
XX method of the invention.
XX
XX Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;
SQ
Query Match 16.2%; Score 81.2; DB 21; Length 37856;
Best Local Similarity 55.2%; Pred. No. 0.00035;
Matches 158; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
QY 211 CAGAGCCGCGCCGCCGCCGCCGACGCCGCCGAGGCCGCCGCCGAGCCCTCCAGGT 270
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 30210 CCTGGCTGCGCGCGCGCGCGCGCTCCGTTGACAGGAGCCGCCGCCGCTCCCGCGC 30151
QY 271 GACCTTGTCCGCGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 30150 GCGCGCGCTTCCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30091
QY 331 GCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 30090 GCGCGCGAGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 30031
QY 391 CACAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 30030 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29971
QY 451 GCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 29970 CCCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29925

RESULT 4
AA199683/c
ID AA199683 standard; DNA: 4403765 BP.
XX
XX AA199683:
XX
XX 15-JAN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX

```


QY 387 TGGACACAGCGGCGCTGAAGCAGCTCCGCGCGCCCGCCGACCGCCGACATGTCG 446
DB 1806 CTCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1865
QY 447 GCGCCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
DB 1866 CTCCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917

RESULT 15

AAZ23778

ID AAZ23778 standard; DNA; 8705 BP.

AC AAZ23778;

DT 14-JAN-2000 (first entry)

DE Vector pshuttle DNA.

KW Antisense; DNA library; identification; multiple cloning site; MCS;
inhibition; ss.

OS Synthetic.

PN WO950457-A1.

PD 07-OCT-1999.

PF 28-MAR-1999; 99WO-US06742.

PR 28-MAR-1998; 98US-0079792.

PR 06-NOV-1998; 98US-0107504.

PA (UTAH) UNIV UTAH RES FOUND.

PI Ruffner DE, Pierce ML, Chen Z;

DR WPI; 1999-610866/52.

PT Production of antisense libraries, used for identifying antisense

PT agents and for identifying target sites for antisense-mediated

PT inhibition of a selected gene -

PS Claim 16; Page 43-50; 63pp; English.

XX This invention describes a novel method for generating an antisense

XX library targeted to a selected RNA transcript. The methods can be used

XX for identifying antisense agents and for identifying target sites for

XX antisense-mediated inhibition of a selected gene. The use of a direct

XX library for target site selection significantly simplifies the screening

XX process, since only very small libraries need be prepared and assayed.

XX This sequence represents the vector pshuttle which is used in the method

XX of the invention.

SQ Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other;

Query Match 14.9%; Score 74.4; DB 20; Length 8705;
Best Local Similarity 48.7%; Pred. No. 0.0043;
Matches 230; Conservative 0; Mismatches 241; Indels 1; Gaps 1;

QY 27 TCCCGGTTTCTTCTGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 86
DB 7312 TCCGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7371
QY 87 GGGGGTGGGCGCGCTGATGAACTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 146
DB 7372 TGTCTGCTGCG 7431
QY 147 TGCAGGTGCGGCG 206
DB 7432 TCTGCG 7491
QY 207 CATTCAGGCG 266

DB 7492 CCGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7551
QY 267 AGGTGACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 326
DB 7552 TGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7610
QY 327 CTGCG 386
DB 7611 CTCCTGCG 7670
QY 387 TGGACACAGCGGCGCTGAAGCAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
DB 7671 CTCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7730
QY 447 GCGCCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
DB 7731 CTCCTGCG 7782

Search completed: October 10, 2002, 15:12:59
Job time : 1204 secs

FILING DATE: 26-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,859
FILING DATE: 29-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/621,443
FILING DATE: 29-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/621,457
FILING DATE: 30-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: G15181A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4405 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 408..2789
US-08-745-880-3

Query Match
Best Local Similarity 8.7%; Score 43.4; DB 2; Length 4405;
Matches 80; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 334 GCGCCAGCGGCTTCCTCGCGTAGAGCGAGGCGCGCGCGATGAAGCGGTGAGCCCG 393
DB 209 CGCGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 150
QY 394 GTGCGCCCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 453
DB 149 GCGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 90
QY 454 TGCGTGGCGGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 474
DB 89 GGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 69

RESULT 4
US-08-480-382-3/c
Sequence 3, Application US/08480382
Patent No. 5986079
GENERAL INFORMATION:
APPLICANT: Bart, Phillip J.
APPLICANT: Brake, Anthony J.
APPLICANT: Kautman, Rhadai J.
APPLICANT: Tekamp-Olson, Patricia
APPLICANT: Wasley, Louise
APPLICANT: Wong, Polly A.
TITLE OF INVENTION: Expression of PACE in Host Cells and
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson & Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,382
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/885,972
FILING DATE: 20-MAY-1992
APPLICATION NUMBER: US 07/621,092
FILING DATE: 26-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,859
FILING DATE: 29-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/621,443
FILING DATE: 29-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/621,457
FILING DATE: 30-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: G15181A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4405 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 408..2789
US-08-480-382-3

Query Match
Best Local Similarity 8.7%; Score 43.4; DB 2; Length 4405;
Matches 80; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 334 GCGCCAGCGGCTTCCTCGCGTAGAGCGAGGCGCGCGCGATGAAGCGGTGAGCCCG 393
DB 209 CGCGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 150
QY 394 GTGCGCCCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 453
DB 149 GCGGCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 90
QY 454 TGCGTGGCGGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 474
DB 89 GGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 69

RESULT 5
US-08-486-343A-1
Sequence 1, Application US/08486343A
Patent No. 6071695
GENERAL INFORMATION:
APPLICANT: OKAYNAK, ENGIN
APPLICANT: OPPERMAN, HERMAN
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING
TITLE OF INVENTION: MORPHOGENIC PROTEIN EXPRESSION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
STREET: INC.
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 07148
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 5970
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-21

Query Match 8.6%; Score 43.2; DB 3; Length 5970;
Best Local Similarity 47.4%; Pred. No. 0.36;
Matches 129; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 219 TCCCGTCATGTTGGTGGGAGTGTGCGCGCCCGAGGCGCGCGGAGGCG 278
DB 5793 TCCCGGCGGTGTGTGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGG 5734
QY 279 AAGGAGCGGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 338
DB 5733 TGAAGTCCCGGTGTCGCCAGGCGGTCCCTGGCGTCCGCGTCCGCGGCGT 5674
QY 339 CAGGCGGTTCCTGCGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 398
DB 5673 CTCGGGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 5614
QY 399 CCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGT 458
DB 5613 CCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5554
QY 459 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490
DB 5553 GTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5522

RESULT 8

US-08-658-136-4
Sequence 4, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LINDER, GREGORY M
APPLICANT: BURR, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMANO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GENA-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 14060 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 135..13040
US-08-658-136-4

Query Match 8.6%; Score 43.2; DB 3; Length 14060;
Best Local Similarity 50.0%; Pred. No. 0.37;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 249 CCGTCCCCGAGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 308
DB 68 CTGAGCTGGCGGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGC 127
QY 309 CCGAGCTTGCTGCTCCCTCCCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 368
DB 128 CCTAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCTGTG 187
QY 369 GCGCGGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 428
DB 188 GCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCGAGGCCCTGACT 247
QY 429 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 464
DB 248 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 283

RESULT 9

US-08-804-227C-7/C
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sulton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:

NAME/KEY: CDS
 LOCATION: 14046..20036
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 20110..31284
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 31329..36071
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36155..41830
 US-08-804-227C-7

Query Match 8.6%; Score 43.2; DB 2; Length 44377;
 Best Local Similarity 48.0%; Pred. No. 0.4;
 Matches 123; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 245 GTCCGCGTGTCCCGAGCGCGCGCGAGGCAAGGAGCGCGCGCGAGC 304
 DB 43302 GTTGGGGGCGCATCCGCGGTGAGCTGGGGCGCGAGCGAGTCCCGCGCGG 43243
 QY 305 GGGCCCGAGCTTGGCTGCTCTCTCGCTCGCCCGAGCGGGTGTGCGGTAGAGGCA 364
 DB 43242 TGGCGCGGTGCGCGCGCGCGCATCTCGAGAGCTGAGCATCGAGCGCGTGAAGGCC 43183
 QY 365 GGGCGCGCGCATGTGAAGCGGTGAGCCCGGTGCGCCCTCGCGCGCGCAAGCGCGTGGC 424
 DB 43182 GGGCGCATCTCGCGCAAGGTGCAACCGCGGTCAACCGTGGGAGCGAGCGCGCGAGCA 43123
 QY 425 GCTGGCGGCGCGGAGCTGCGCTGCGCTGCGCGAGCGAGCGCGCGCTGGAGT 484
 DB 43122 CTCGTGGGCGGTGTGGTGGCGCGCGCGCGAGGGCGCTGGGGCGGTACGGGCTGAGCA 43063
 QY 485 GCTCCGCGAGCGCGCGC 500
 DB 43062 CCGGGCGCGCGAGCGC 43047

RESULT 10
 US-08-804-198-1/c
 Sequence 1, Application US/08804198
 Patent No. 5945320
 GENERAL INFORMATION:
 APPLICANT: Burgett, Stanley G.
 APPLICANT: Kuhnstoss, Stuart A.
 APPLICANT: Rao, Nagendra R.
 APPLICANT: Richardson, Mark A.
 APPLICANT: Rostek, Paul R., Jr.
 TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PAUL R. CANTRELL 1138
 STREET: LILLY CORPORATE CENTER
 CITY: INDIANAPOLIS
 STATE: IN
 COUNTRY: USA
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Macintosh
 OPERATING SYSTEM: Macintosh 7.0
 SOFTWARE: Microsoft Word 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/804.198
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CANTRELL, PAUL R.
 REGISTRATION NUMBER: 36,470
 REFERENCE/DOCKET NUMBER: P9113
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-3885
 INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 44377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 350..14002
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 14046..20036
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 20110..31284
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 31329..36071
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36155..41830
 US-08-804-198-1

Query Match 8.6%; Score 43.2; DB 2; Length 44377;
 Best Local Similarity 48.0%; Pred. No. 0.4;
 Matches 123; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 245 GTCCGCGTGTCCCGAGCGCGCGCGAGGCAAGGAGCGCGCGCGAGC 304
 DB 43302 GTTGGGGGCGCATCCGCGGTGAGCTGGGGCGCGAGCGAGTCCCGCGCGG 43243
 QY 305 GGGCCCGAGCTTGGCTGCTCTCTCGCTCGCCCGAGCGGGTGTGCGGTAGAGGCA 364
 DB 43242 TGGCGCGGTGCGCGCGCGCGCATCTCGAGAGCTGAGCATGCTGAGCGCGTGAAGGCC 43183
 QY 365 GGGCGCGCGCATGTGAAGCGGTGAGCCCGGTGCGCCCTCGCGCGCGCAAGCGCGTGGC 424
 DB 43182 GGGCGCATCTCGCGCAAGGTGCAACCGCGGTCAACCGTGGGAGCGAGCGCGCGAGCA 43123
 QY 425 GCTGGCGGCGCGGAGCTGCGCTGCGCTGCGCGAGCGAGCGCGCGCTGGAGT 484
 DB 43122 CTCGTGGGCGGTGTGGTGGCGCGCGCGCGAGGGCGCTGGGGCGGTACGGGCTGAGCA 43063
 QY 485 GCTCCGCGAGCGCGCGC 500
 DB 43062 CCGGGCGCGCGAGCGC 43047

RESULT 11
 US-09-428-517-1/c
 Sequence 1, Application US/09428517
 Patent No. 6251636
 GENERAL INFORMATION:
 APPLICANT: Betlach, Mary C.
 APPLICANT: Shah, Sanjay Krishnakant
 APPLICANT: McDaniel, Robert
 APPLICANT: Tang, Li
 TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
 FILE REFERENCE: 30062-20029.00
 CURRENT APPLICATION NUMBER: US/09/428.517
 EARLIER FILING DATE: 1999-10-28
 EARLIER APPLICATION NUMBER: 60/120.254
 EARLIER FILING DATE: 1999-02-16
 EARLIER APPLICATION NUMBER: 60/106.100
 EARLIER FILING DATE: 1998-10-29
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 50937
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA


```

/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone="IMAGE:5285103"
/clone_1lb="NH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notice="Organ: brain; Vector: pBluescriptII (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcaca
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTNN-3',
size selected for average insert size 2.5 kb and

```

normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHRI, National Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 135 a 268 c 320 g 119 t 1 others

Query Match 90.0%; Score 450.2; DB 10; Length 843;
Best Local Similarity 98.6%; Pred. No. 2.8e-66;
Matches 486; Conservative 0; Mismatches 3; Indels 4; Gaps 3;

QY 10 GCGCGGTTGTGATGATACCGGAGTGGGATCCCGGGCTAGGGAGCGGCGG -C 67
DB 23 GCGCGGTTGTGATGATACCGGAGTGGGATCCCGGGCTAGGGAGCGGCGGCG 82
QY 68 CCGATGGGCTGTGGGAGCTCCGAAGGAGTACTAGGACACCGGGGGGCTACTTT 127
DB 83 GCGATGGGCTGTGGGAGCTCCGAAGGAGTACTAGGACACCGGGGGGCTACTTT 142
QY 128 TCTTCCGCTGCTTTGCTTTTTCCTTTGGGCTGGGCTGAGTGCCTCCACTGAGC 187
DB 143 TCTTCCGCTGCTTTGCTTTTTCCTTTGGGCTGGGCTGAGTGCCTCCACTGAGC 201
QY 188 AAGATTTCCCTGCTAAACCCAGAGGACCCCTCCGCAATTGTTGGGCTGGGAGTGC 247
DB 202 AAGATTTCCCTGCTAAACCCAGAGGACCCCTCCGCAATTG-TGGGCTGGGAGTGC 260
QY 248 GCGGTCCTCCGAGCGCGCGCGGAGGCAAGGAGCGAGCGCGCGGAGCGGCG 307
DB 261 GCGGTCCTCCGAGCGCGCGCGGAGGCAAGGAGCGAGCGCGCGGAGCGGCG 320
QY 308 CCGGAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 367
DB 321 CCGGAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 380
QY 368 CCGGCGGATGAAGGCGGTGAGCCCGGCTGCGCCCTGCGGCGCAAGGCGCGT 427
DB 381 CCGGCGGATGAAGGCGGTGAGCCCGGCTGCGCCCTGCGGCGCAAGGCGCGT 440
QY 428 GCGGCGGCGGAGCGGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487
DB 441 GCGGCGGCGGAGCGGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
QY 488 CCGGAGCGCGGCG 500
DB 501 CCGGAGCGCGGCG 513

RESULT 2
BI552992 570 bp mRNA linear EST 05-SEP-2001
LOCUS 603193628F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5264694 5',
DEFINITION mRNA sequence.
ACCESSION BI552992
VERSION BI552992.1 GI:15440304
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 570)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: egapbs-remail.nih.gov
Tissue Procurement: MIKLOS Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHRI), Shihaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM1667 row: h column: 07

High quality sequence stop: 543.

Location/Qualifiers

1..570

/organism="Homo sapiens"

/db.xref="taxon:9606"

/clone="IMAGE:5264694"

/clone_1ib="NIH_MGC_95"

/tissue_type="hippocampus"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescript (modified

pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag

; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',

size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIMH/NHRI, National

Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 72 a 174 c 239 g 85 t

Query Match 88.8%; Score 444; DB 10; Length 570;
Best Local Similarity 98.3%; Pred. No. 3.3e-65;
Matches 470; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 10 GCGCGGTTGTGATGATACCGGAGTGGGATCCCGGGCTAGGGAGCGGCGG -C 67
DB 22 GCGCGGTTGTGATGATACCGGAGTGGGATCCCGGGCTAGGGAGCGGCGGCG 81
QY 68 CCGATGGGCTGTGGGAGCTCCGAAGGAGTACTAGGACACCGGGGGGCTACTTT 127
DB 82 GCGATGGGCTGTGGGAGCTCCGAAGGAGTACTAGGACACCGGGGGGCTACTTT 141
QY 128 TCTTCCGCTGCTTTGCTTTTTCCTTTGGGCTGGGCTGAGTGCCTCCACTGAGC 187
DB 142 TCTTCCGCTGCTTTGCTTTTTCCTTTGGGCTGGGCTGAGTGCCTCCACTGAGC 200
QY 188 AAGATTTCCCTGCTAAACCCAGAGGACCCCTCCGCAATTGTTGGGCTGGGAGTGC 247
DB 201 AAGATTTCCCTGCTAAACCCAGAGGACCCCTCCGCAATTGTTGGGCTGGGAGTGC 260
QY 248 GCGGTCCTCCGAGCGCGCGCGGAGGCAAGGAGCGAGCGCGCGGAGCGGCG 307
DB 261 GCGGTCCTCCGAGCGCGCGCGGAGGCAAGGAGCGAGCGCGCGGAGCGGCG 320
QY 308 CCGGAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 367
DB 321 CCGGAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 380
QY 368 CCGGCGGATGAAGGCGGTGAGCCCGGCTGCGCCCTGCGGCGCAAGGCGCGT 427
DB 381 CCGGCGGATGAAGGCGGTGAGCCCGGCTGCGCCCTGCGGCGCAAGGCGCGT 440
QY 428 GCGGCGGCGGAGCGGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
DB 441 GCGGCGGCGGAGCGGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 498

RESULT 3
BG337080 911 bp mRNA linear EST 27-FEB-2001
LOCUS 602434352F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4552357 5',
DEFINITION mRNA sequence.
ACCESSION BG337080
VERSION BG337080.1 GI:13143518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information

This 5' resequenced clone has no previous 5' data to verify this
new read against
Insert Length: 644 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 466.

FEATURES

source

1. 522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1500639"
/clone_lib="NCI_CGAP_K1d3"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 79 a 200 c 172 g 70 t 1 others
ORIGIN

Query Match 73.7%; Score 368.4; DB 9; Length 522;
Best Local Similarity 99.5%; Pred. No. 1.5e-52;

Matches 369; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 130 TTCGGGCTTTGCTTTTCTTTTCTTTGCGCTGCGCTGAGTGTCCGACCA 189
DB 522 TTCGGGCTTTGCTTTTCTTTTCTTTGCGCTGCGCTGAGTGTCCGACCA 463
OY 190 AGATTCCTCGTAACACAGAGACACCTCCGTCATTTGTTGGGCTCGGAGTGTCCG 249
DB 462 AGATTCCTCGTAACATCAGAGACACCTCCGTCATTTGTTGGGCTCGGAGTGTCCG 403
OY 250 GGTGCCCCGAGCGCGCGCGCGGAGGCAAGGAGCGGACCGCGCGAGCGGGGCC 309
DB 402 GGTGCCCCGAGCGCGCGCGCGGAGGCAAGGAGCGGACCGCGCGAGCGGGGCC 343
OY 310 CGGACCTTGCCCTCCCTCGCTGCGCCACAGGGGTTGCTGCGTAGAGCGACAGGCG 369
DB 342 CGGACCTTGCCCTCGCTGCGCCACAGGGGTTGCTGCGTAGAGCGAGGGCG 283
OY 370 CGCGGATGAAGCGGATGAGCCCGGTCGCGCCCTCGGGCGCAAGGCGCGCTCGGGCTGC 429
DB 282 CGCGGATGAAGCGGATGAGCCCGGTCGCGCCCTCGGGCGCAAGGCGCGCTCGGGCTGC 223
OY 430 GCGCGGGGAGCTGGCGCTGCGCTGCGCGGAGCAGCGGCAAGCTGGGTGCTCG 489
DB 222 GCGCGGGGAGCTGGCGCTGCGCTGCGCGGAGCAGCGGCAAGCTGGGTGCTCG 163
OY 490 GCAGCGCGGGC 500
DB 162 GCAGCGCGGGC 152

RESULT 7
LOCUS B1546228 826 bp mRNA linear EST 05-SEP-2001
DEFINITION 603188614F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260155 5',
ACCESSION B1546228
VERSION B1546228
KEYWORDS mRNA sequence.
SOURCE EST.
ORGANISM human.
human sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 826)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLML at:
<http://image.lnl.gov>
Plate: L1AM1655 row: k column: 04
High quality sequence stop: 821.

FEATURES

source

1. 826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:5260155"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pBluescript (modified
pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 128 a 261 c 307 g 130 t
ORIGIN

Query Match 73.6%; Score 367.8; DB 10; Length 826;
Best Local Similarity 96.4%; Pred. No. 1.7e-52;

Matches 483; Conservative 0; Mismatches 7; Indels 11; Gaps 10;

OY 10 GGGGGGTTGTGATGATCCGAGTGGGATGATCCCGGCTAGGAGCGCGGCG--C 67
DB 23 GGGCGGTTGTGATGATCCGAGTGGGATGATCCCGGCTAGGAGCGCGGCGCG 82
OY 68 CCGATCGGCTTAGTGGAGCTCCGAAGGAGTACTAGACACCGGGGTGCTACTTT 127
DB 83 GCGATCGGCTTAGTGGAGCTCCGAAGGAGTACTAGACACCGGGGTGCTACTTT 142
OY 128 TCTTCCGGTCTTTTCTTTTCTTTTCTTTGCGCTCGGCTGAGTGTGCCCACTGAGC 187
DB 143 TCTTCCGGTCTTTTCTTTTCTTTTCTTTGCGCTCGGCTGAGTGTGCCCACTGAGC 201
OY 188 AAAGATTCCCTCGTAACACAGAGACACCTCCCGTCATTTGTTGGGCTCGGAGATGTC 247
DB 202 AAAGATTCCCTCGTAACACAGAGACACCTCCCGTCATTTGTTGGGCTCGGAGATGTC 261
OY 248 GCGGTGCCCA--GCGGCGCGGCGGAGCAAGGAGCGGAGCGGCG--CGCGAGCG 305
DB 262 GCGGTGCCCAAGTGGCGCGCGGCGGAGCAAGGAGCGGAGCGGAGTGGCGG 321
OY 306 GCGCGGAGCTTGGCTCTCCCTCGCTCGGCCCA--GCGGGTTGCGTCCG--GTAGAGCG 363
DB 322 GCGCGGAGCTTGGCTCTCCCTCGCTCGGCCCAAGGAGCGGAGTGGCGGAGGCGC 381
OY 364 AGGCGCG--GCGGATGAAGC--GGTGAAGCCCGGTGCGCCCTCG--GGCCGCAAGGCGCG 420
DB 382 AGGCGCGCTGCGCATGAAGCTGTGAGCCCGGTGCGCCCTCGTGGCCACAAAGCGCG 441
OY 421 TCGGCTGCGGCG--GCGGGAAGTGGCGCTGCGCTGCGCGAGCAGCGCACAGCT 479
DB 442 TCGGCTGCGGCGGCGGAGTGGCGCTGCGCTGCGCGAGCAGCGCACAGCT 501
OY 480 GGTGCTCGGCAAGCGCGCG 500
|||||

Db 502 GGGTGGCTCCGACGCCGCGC 522

RESULT 8
LOCUS BE327284
DEFINITION BE327284 472 bp mRNA linear EST 14-JUL-2000
hw10a02.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3182474 3' similar to contatus MER22.b3 TARI repetitive element ;, mRNA sequence.
BE327284
VERSION BE327284.1 GI:9201060
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400p from Glibcoo
High quality sequence stop: 421.
Location/Qualifiers
1..472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3182474"
/clone_lib="NCI-CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 56 a 158 c 186 g 72 t

ORIGIN

Query Match 71.6%; Score 358; DB 10; Length 472;
Best Local Similarity 99.7%; Pred. No. 8.2e-51;
Matches 369; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 132 CCGGTGCTTGGC-TTTTTTTCCTTGGCTGGCTGAGTGTGCCCACTGAGCAAA 190
|||||
Db 1 CCGGTGCTTGGCTTTTTTTTCTTGGCTGGCTGAGTGTGCCCACTGAGCAAA 60
|||||

QY 191 GATTCCCTGTAAACCCAGAGCACCCTCCGTCATTTGTTGGCTGGGAGTGTCCG 250
|||||
Db 61 GATTCCCTGTAAACCCAGAGCACCCTCCGTCATTTGTTGGCTGGGAGTGTCCG 120
|||||

QY 251 GTGCCCGGAGCGCGCGGCGGAGGCAAGGAGCGGAGCCGCGGAGCGGCGCC 310
|||||
Db 121 GTGCCCGGAGCGCGCGGCGGAGGCAAGGAGCGGAGCCGCGGAGCGGCGCC 180
|||||

QY 311 GGAGCTTGCTGCTCCGCTCGCTCGCCCAAGCGGGTTCGCTCCGCTAGAGCGGAGCGGCG 370
|||||
Db 181 GGAGCTTGCTGCTCCGCTCGCTCGCCCAAGCGGGTTCGCTCGCTAGAGCGGAGCGGCG 240
|||||

QY 371 GCGGATGAAGCGCGGTGAGCCCGGTGCGCCCTCGGCGCCAGGCGCCGTGGCTGC 430
|||||
Db 241 GCGGATGAAGCGCGGTGAGCCCGGTGCGCCCTCGGCGCCAGGCGCCGTGGCTGC 300
|||||

QY 431 GCGGCGGGAGCTGGCGCTGGCTGCTCCGTGGCCGACACAGCCGAGCTGGCTCGC 490
|||||
Db 301 GCGGCGGGAGCTGGCGCTGGCTGCTCCGTGGCCGACACAGCCGAGCTGGCTCGC 360
|||||

QY 491 CAGCCGCGGC 500
|||||
Db 361 CAGCCGCGGC 370
|||||

RESULT 9
LOCUS B1598458
DEFINITION B1598458 817 bp mRNA linear EST 07-SEP-2001
603246969F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5294681 5', mRNA sequence.
B1598458
VERSION B1598458.1 GI:15491397
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 817)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Niklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshitaki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1745 row: 1 column: 18
High quality sequence stop: 386.
Location/Qualifiers
1..817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5294681"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pBluescript (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTCTTTTCTTTTCTTTT-3', size-selected for average insert size 2.3 kb and normalized to R0F 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 158 a 223 c 316 g 120 t

ORIGIN

Query Match 59.8%; Score 299.2; DB 10; Length 817;
Best Local Similarity 92.3%; Pred. No. 5.2e-41;
Matches 349; Conservative 0; Mismatches 23; Indels 6; Gaps 3;

QY 10 GCGCGGTTGTGAGTACCGGAGTGGGATGCCGGCTAGGGGAGCGGCGC--C 67
|||||
Db 20 GCGCGGTTGTGAGTACCGGAGTGGGATGCCGGCTAGGGGAGCGGCGC 79
|||||

QY 68 CGGATCGGCTTAGTGGAGCTCCGAAGGAGTACTAGACACCCGGGTGGCTACTTT 127
|||||
Db 80 CGGATCGGCTTAGTGGAGCTCCGAAGGAGTACTAGACACCCGGGTGGCTACTTT 139
|||||

QY	128	TCCTCCGAGCTCTTGGTTTTTTTTCCTTGGGCTCGGGCTGATGTCGCCACTGAGC	187
QY	128	TCCTCCGAGCTCTTGGTTTTTTTTCCTTGGGCTCGGGCTGATGTCGCCACTGAGC	187
Db	140	TCCTCCGAGCTCTTGGTTTTTTTTCCTTGGGCTCGGGCTGATGTCGCCACTGAGC	198
QY	188	AAAGATTCCTCTGTAAAAACCAGAGCAGCCCTCCGTCAAATTGTTGGGCTCGGGAGTGTCT	247
Db	199	AAAGATTCCTCTGTAAAAACCAGAGCAGCCCTCCGTCAAATTGTTGGGCTCGGGAGTGTCT	258
QY	248	GCGGTGCGCCGAGAGCGCGCGGGCGGAGGCAAAAGGAGGAGCAGCGGCGCGGAGCGAGG	307
Db	259	GCGGTGCGCCGAGAGCGCGCGGGCGGAGGCAAAAGGAGGAGCAGCGGCGCGGAGCGAGG	318
QY	308	CCTCGAGCTTCCCTGCGCTCTGCTGCGCCCGAGCGGGCTTCGCTCGGTAGAGCGCA	364
Db	319	CCTCGAGCTTCCCTGCGCTCTGCTGCGCCCGAGCGGGCTTCGCTCGGTAGAGCGCG	378
QY	365	GCGGCGCGCGGAGTGAAGG	382
Db	379	AGGGGCGGCGCGGAGTGAAGG	396

RESULT 10	LOCUS	DEFINITION
AM028893	AM028893	310 bp mRNA linear EST 20-OCT-2000
	W97G12.X1 NCI-CGAP_Gas4	Homo sapiens CDNA clone IMAGE:2542102 3'
	mRNA sequence.	

ACCESSION	AW028893
VERSION	AW028893.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 310)
AUTHORS	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D.

JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remall.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert length: 298 Std Error: 0.00 Seq primer: -40up from Gibco High quality sequence stop: 302. Location/Qualifiers I. .310
FEATURES	
SOURCE	

BASE COUNT	ORIGIN
39 a	102 c 110 g 58 t 1 others

Query Match	58.6%	Score 293.2	DB 9	length 310
Best Local Similarity	98.4%	Pred. No. 66-40		
Matches 306	Conservative 0	Mismatches 4	Indels 1	Gaps 1
<p>QY 99 GTCAGCAGACACCGGGGGCTACTTTTCCTCCGTCCTTTTCTTTTTCCTTT 158</p> <p> </p>				

Db	1	GTGACTAGACACCCGGGTGGGCTACTTTTCTTCCGGTCTTGC-TTTTTTTCCTT	59
QY	159	GGGCTGGGGCTGAATGTGCCCACTAGACAAATATCCCTGTAACCAACGAGGGACCC	218
Db	60	GGGCTCGGGGTGAATGTGCCCACTAGACAAATATCCCTGTAACCAACGAGGCCACC	119
QY	219	TCCCGTCAATTTGTTGGGCTCGGGAGTGTCCGGTGCCTCGAGCGCGCGGGCGGAGGC	278
Db	120	TCCCGTCAATTTGTTGGGCTCGGGAGTGTCCCGGTGCCCAAGCGCGCGGGCGGAGGC	179
QY	279	AAAGGAGCGGAGCGCGGCGCGGGGCGCGGAGCTTCGTCGCTCCCTCGCTGCGCC	338
Db	180	AAAGGAGCGGAGCGCGGCGCGGGGCGCGGAGCTTTCGCTCCCTCGCTGCGCC	239
QY	339	CAGGCGGTTTCGCTGCGGTAGAGCGCGAGGGCGCGCGCATGAAGCGGTGAGCCCGGTGCG	398
Db	240	CAGGCGGTTTCGCTGCGCGAGGGCGCGAGCGCGCATGAAGCGGTGAGCCCGGTGCG	299
QY	399	CGCCCTCGGGCC 409	
Db	300	CGCCCTCGGGCC 310	

RESULT 11			
BF711180/c			
LOCUS	BF711180	455 bp	linear
DEFINITION	M-P-A11-nrc-h-10-0-UI s1 M1-P-A11 Sns scrofa cDNA clone		
	M1-P-A11-nrc-h-10-0-UI 3', mRNA sequence.		

ACCESSION	BF711180	GI:12010657
VERSION	BF711180.1	
KEYWORDS	EST.	
SOURCE	plg.	

ORGANISM
SUS SCROFA
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE
AUTHORS
TITLE
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
Contact: Tugale CK
Laboratory, Department of Animal Science

COMMENT

Contact: Tugge CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildeer Hall, Ames, IA 50011-3150, USA
Tel: 5152844252
Fax: 5152942401
Email: cktugge@iastate.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares lab, University of Iowa Esr sequencing: M.B. Soares Lab,
University of Iowa Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.

```

FEATURES
source
Location/Qualifiers
1..455
/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-A11-nrc-h-10-0-01"
/clone_lib="MI-P-A11"
/lab_host="DHI08 (Life Technologies)"
/notes="vector: pUT3D-pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: EcoRI. The MI-P-A10
library is normalized library derived from the MI-P-A10
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigdb.genome.jaxstate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_SEQ=None found"

```


Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401

Email: cktugle@iastate.edu
Oligo-dt track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
University of Iowa Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.

FEATURES

Location/Qualifiers
1..440
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="MI-P-AV0-nev-h-05-0-UI"
/clone_1ib="MI-P-AV0"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: EcoRI; The MI-P-AV0
library is derived from placenta. For a detailed
description of the library from which this clone was
derived, please visit our web site at
http://piglet.genome.iastate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_SEQ=None found"

BASE COUNT 62 a 168 c 145 g 65 t

ORIGIN

Query Match 49.7%; Score 248.4; DB 10; Length 440;

Best Local Similarity 79.0%; Pred. No. 1.8e-32;

Matches 346; Conservative 0; Mismatches 61; Indels 31; Gaps 3;

80 AGCGAGCTCCGAGGAGTACTAGACACCGGGGCTACTTTCTCCGGTCT 139
|||||
428 AGCGAGGAGTCCGAGGAGTACTAGACACCGGGGCTACTTTCTCCGGTCT 369
|||||
140 TTTGCTTTTCTCTTGGCTCGGGCTGAGTGTGCCCACTGAGCAAAAGATTCCCTC 199
|||||
368 ATTGATTTTCTTCCAAATGATGTCGCG-----GCCACAGACAGAGATCCCTC 319
|||||
200 GTAAACCCGAGGAGCCCTCCGTCATTTGTTGGCTCGGAGTGTGCGG----- 250
|||||
318 GCAAAAGCGAGGAGCCCTCCGTCATTTGTTGGCTCGGAGTGTGCGG----- 259
|||||
251 -----GTGCCGAGCGCGCGCGGCGGAGCAAAAGGAGCGAGCGCGC 299
|||||
258 GTGCCCGAGCGGTCGACCGAGCGGCGGCGGAGCAAAAGGAGCGAGCGCGC 199
|||||
300 -GAGAGGGGCGGAGCTTTCCTCTCCCTCGCTCGCCAGGCGGTTGCGCGTAG 358
|||||
198 GGGTGGGGGCGGAGCTTTCCTCTCCCTCGCTCGCCAGGCGGTTGCGCGTAG 139
|||||
359 AGCGAGGCGCGCGATGAAGCGGCTGAGCCGCTGCGCCCTCGGGCGCAAGCGC 418
|||||
138 AAGCGAGGCGCGCGATGAAGCGGCTGAGCCGCTGCGCCCTCGGGCGCTAAGCGC 79
|||||
419 CGTGGGCTCGGCGGCGGAGCTGCGCTGCGCTGCGCCAGAGCGGCAAGCC 478
|||||
78 CGTGGGCTCGGCGGCGGAGCTGCGCGCGCTGCGCTGCGCCAGAGCGCAAGCC 19
|||||
479 TGGTGGCTCGGCGGCGG 496
|||||
18 TGGCGGCTCGGCGTCCG 1

RESULT 14

BI398600/c 385 bp mRNA linear EST 14-AUG-2001

LOCUS MI-P-AV1-npu-e-10-0-UI.s1 MI-P-AV1 Sus scrofa cDNA clone
DEFINITION MI-P-AV1-npu-e-10-0-UI 3', mRNA sequence.

ACCESSION BI398600 GI:15177661
VERSION BI398600.1
KEYWORDS EST.

SOURCE

ORGANISM

PIG.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Tugle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktugle@iastate.edu
Oligo-dt track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
University of Iowa Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.

FEATURES

source

Location/Qualifiers
1..385
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-AV1-npu-e-10-0-UI"
/clone_1ib="MI-P-AV1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: EcoRI; The MI-P-AV1
library is normalized library derived from the MI-P-AV0
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://piglet.genome.iastate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_SEQ=None found"

BASE COUNT 59 a 146 c 136 g 44 t

ORIGIN

Query Match 49.1%; Score 245.4; DB 10; Length 385;

Best Local Similarity 83.0%; Pred. No. 5.8e-32;

Matches 328; Conservative 0; Mismatches 36; Indels 31; Gaps 3;

124 CTTTCTTCCGGTGTCTTTCTTTTCTTGGCTGGGCTGAGTGTGCGCCACT 183
|||||
385 CTTTCTTCCGGTGTCTTTCTTTTCTTGGCTGGGCTGAGTGTGCGCCACT 336
|||||
184 GAGCAAGATTCCTCGTAAACCGAGACCGCTCCGTCATTTGTTGGCTGGGAG 243
|||||
335 GAGCAGAGATCCCTCGCAAAAGCAGACCACTCCGTCATTTGTTGGCTGGGAGT 276
|||||
244 TGTGCG-----GTGCCGAGGCGCGCGGCGGCGGAGGCAAGG 283
|||||
275 CGCGGAGCTTCCGCGGTGCGCCCGAGCTGAGCGCAACCGGCGGAGGCAAGG 216
|||||
284 GAGCGAGCGCGCGCGC-GAGCGGCGCGGAGCTTCCGCTCCCTCGTGGCCAGC 342
|||||
215 GAGCGAGCGCGCGCGCGGCGGCGGAGCTTCCGCTCCCTCGTGGCCAGC 156
|||||
343 GGGTTCGCTCGGTAGAGCGAGGCGCGCGGATGAAGCGGCTGAGCCCGTGGCCCC 402
|||||
155 GGGTTCGCTCGGTAGAGCGAGGCGCGCGGATGAAGCGGCTGAGCCCGTGGCCCC 96
|||||
403 TGGGCGCGCAAGCGCGCTGGGCTGCGGCGGCGGAGGAGTGGCGCTGGCGTGGCC 462

DE	Regulator of neuronal proliferation RNP-1 partial DNA.
XX	Regulator of neuronal proliferation; neuron; RNP-1; ss.
KW	Homo sapiens.
OS	MO9516774-A1.
PN	22-JUN-1995.
XX	19-DEC-1994; 94MO-US14614.
PD	08-SEP-1994; 94US-0301416.
PR	17-DEC-1993; 93US-0169522.
XX	(SPIN-) SPINAL CORD SOC.
PA	Neuman T, Nornes HO, Suda K;
PI	WPI: 1995-231568/30.
DR	A vector contg. a nucleic acid encoding an E2F and/or E1A regulator
XX	- is used to induce DNA synthesis in differentiated neurons and
PF	replace neurons after injury or neurodegenerative diseases.
PS	Disclosure; Page 30; 60pp; English.
CC	This sequence corresponds to RNP-1 which has no significant
CC	homology to other characterized sequences.
SQ	Sequence 1309 BP; 305 A; 339 C; 381 G; 284 T; 0 other;
Query Match	23.4%; Score 116.8; DB 16; Length 1309;
Best Local Similarity	61.88; Pred. No. 1.4e-15;
Matches 220; Conservative	0; Mismatches 132; Indels 4; Gaps 2
DG	1 TTTTCTTCCTTGCGGTGGCGTGGCTAGCTGCACACTGACCAAAATTCCTGATAA 204
OY	1 TTTTCTTCCTTGCTTTGCTTGGCGCATGTGCTCACAABAAAAGAAAAAACAAAA 60
DG	205 ACCGAGAGCACCCCTCCCTCATTTGTGGGCTGGGAGTGCAGCTGCGGTGCCCGGAAGG 264
OY	265 CGGGAGCCGAGAGGCAAAAGAGAGCGAGCCGCGGAGAGGAGCCCGGATCTGCGTCGC 324
DG	118 CTGGAGACCGCGGGGAGTCTGAGACCAGCCGAGACGAGAGAGGAGCAAAAGCAATTGGG 177
OY	325 TCCCTCCCTGCGCCCCAGCGGATTCGCTCGTAGAGCGCGAGGCGCGCGGATGAGAAGCG 384
DG	178 GCGGGAGACCGAGACCTGCTGCTGCTGCTGCTGCTCCAGACAGCGCGGATGAGAAGCG 237
OY	385 GTGAGACCCGGTGGCGCCCTCGGAGCCGAGAGGCGCGCTGGGAGCTGGCGGAGAGCTG 444
DG	238 GTGAGACCCGGTGGCG--CCCTCGGAGGTCAAGAGGCGCGTGGGCGTGGCGGCGGAGCTG 296
OY	445 GCGCTGCGCTGCTGGCCGAGACAGGCGACACAGCTGGGTGGCTCGCGACGCGCGCG 500
DG	297 GCGGTAGCGTGGCTGGCGAGACAGGCGACACAGCTGGGTGGCTCGCGACGCGCGCG 352
RESULT 4	
ID	AAD25519 standard; DNA; 154746 BP.
XX	AAD25519;
XX	26-MAR-2002 (first entry)
DC	Human herpesvirus 2 complete DNA genome.
XX	Human herpesvirus 2; cytosolic; cancer; immunosuppressive; virucide;
KW	antibacterial; fungicide; protozoacide; antitubercular; antiinflammatory;

[illegible]

QY 491 CAGCCGG 498
 DB 1420 GAGCTCCG 1427

RESULT 5
 AAD25519/c
 ID AAD25519 standard; DNA: 154746 BP.
 AC AAD25519;
 DT 26-MAR-2002 (first entry)
 DE Human herpesvirus 2 complete DNA genome.
 KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
 KW antibacterial; fungicide; protozoacide; antitubercular; antiinflammatory;
 KW antiallergic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
 KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
 KW vasculitis; ds.
 XX
 XX Human herpesvirus 2.
 XX W0200176643-A1.
 XX 18-OCT-2001.
 XX 06-APR-2001: 2001WO-US11372.
 XX 07-APR-2000: 2000US-195680P.
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 XX Orson FM, Kinsey BM, Bhogal BS;
 XX WPI: 2002-066308/09.
 XX Composition for oral delivery of vaccines, comprises expression vector
 PT containing antigenic genomic sequence, bound to aggregated
 PT protein-polycationic polymer conjugate or suspension
 XX
 XX Disclosure: Page 90-132; 145pp; English.

The invention relates to a composition comprising an expression vector
 CC bound to an aggregated protein-polycationic polymer conjugate or
 CC suspension. The expression vector contains a promoter polynucleotide
 CC sequence operatively linked to a polynucleotide sequence encoding an
 CC antigen which is a fragment of a gene or genome associated with an
 CC infectious disease, cancer and autoimmune disease such as rheumatoid
 CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
 CC consisting of bacterium, fungus, protozoa and virus such as human
 CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
 CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
 CC optionally comprising a nucleotide sequence encoding a cytokine (or a
 CC cytokine expression vector), is useful for inducing an immune response
 CC (systemic and/or mucosal) in an organism. The cytokine expression vector
 CC contains a sequence for granulocyte macrophage-colony stimulating factor
 CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
 CC the antigen and the cytokine are under transcriptional control of same or
 CC different promoter polynucleotide sequences. The expression vector, as a
 CC DNA vaccine is useful for treating a condition in an organism. The
 CC present sequence is human herpesvirus 2 complete DNA genome related
 CC to the invention.

Sequence 154746 BP: 23003 A; 54218 C; 54701 G; 22824 T; 0 other:
 SO

Query Match 10.3%; Score 51.6; DB 24; Length 154746;
 Best Local Similarity 50.3%; Pred. No. 0.051;
 Matches 155; Conservative 0; Mismatches 149; Indels 4; Gaps 1;
 OY 191 GATTCCTCCGTAACCCAGCAGCCCTCCGCTCAATTTGGCGCCGAGATGTCGCG 250
 DB 126126 GTGCGCTCTGGAGACGCGCCGCTGCGCCACAGGGGCTCTTGCTCCGACGAGCGAC 126067

QY 251 GTGCCCCAGACGCGCGCGCCGAGCAAAAGGAGCGAGCCGCGGACGAGGCGCC 310
 DB 126066 CGACCGGACCGGTTCCGCGCCGCGCGCGCGCGAGCGGTCTATCCGACCTTCCT 125007

OY 311 GGAGCTTCGCGCTCCCTGCGCCGCCAGCGGCTGCTGCGTAAGGAGCGAGGCGC 370
 DB 126006 GAGCCCGAGCGCCGAGCTCGCGCCGAGCCGAGCCCGGAGCAAGAGCGCGGAC 125947

OY 371 GCGGATGAAGGCGGTGAGCCGAGTCCGCGCCCTCGAGCGCGCTGCGGCTGCG 430
 DB 125946 GCGGAGAGAGAGAGAGCGCGCGCGCGGCTCTCCGCGCGCGCGCGCGG 125387

OY 431 GCGGCGGAGAGCTGCGCTGCGCTGCGTGCAGACGCGCCAGACGCTGCGGTCGCG 490
 DB 125886 CCGTTCGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 125831

QY 491 CAGCCGG 498
 DB 125830 GAGCTCCG 125823

RESULT 6
 AAX53491
 ID AAX53491 standard; DNA: 114955 BP.
 AC AAX53491;
 DT 05-JUL-1999 (first entry)
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW pulmonary obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 XX
 XX Synthetic.
 XX W09913886-A1.
 XX 25-MAR-1999.
 XX 17-SEP-1998: 98WO-US19419.
 XX 09-JUN-1998: 98US-0093972.
 XX 17-SEP-1997: 97US-0059160.
 XX (UYEC-) UNIV EAST CAROLINA.
 XX Nyce JW;
 XX WPI: 1999-229400/19.
 XX Antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 PT
 XX Disclosure: Page 37; 120pp; English.

The specification describes antisense oligonucleotides (AAX53491-
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, coding and
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of RNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences AAX53491-74. These multiple target


```
FT /standard_name= "Single nucleotide polymorphism"
FT replace(33840,A)
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT 33881..35300
FT /tag= u
FT /number= 6
FT /cons_splice=(5'site:NO,3'site:NO)
FT replace(33922,G)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT replace(35202,G)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT 35301..35343
FT /tag= x
FT /number= 7
FT 35344..37315
FT /tag= y
FT /number= 7
FT replace(35497,G)
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT replace(37287,A)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT replace(37296,A)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT 37316..37427
FT /tag= ac
FT /number= 8
FT 37428..38381
FT /tag= ad
FT /number= 8
FT /cons_splice=(5'site:YES,3'site:NO)
FT replace(37545,A)
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT 38382..38517
FT /tag= af
FT /number= 9
FT replace(38397,T)
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT 38518..40495
FT /tag= ah
FT /number= 9
FT replace(38622,C)
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT 40496..40562
FT /tag= aj
FT /number= 10
FT replace(40516,C)
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT replace(40521,T)
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT 40563..40748
FT /tag= am
FT /number= 10
FT 40749..40861
FT /tag= an
FT /number= 11
FT replace(40841,A)
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT replace(40875,C)
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism"
XX
```

```
PN WO200190125-A2.
XX
XX 29-NOV-2001.
XX
XX 24-MAY-2001; 2001WO-US16909.
XX
XX 24-MAY-2000; 2000US-206664P.
XX
XX (GENA-) GENAISSANCE PHARM INC.
XX
XX Chew A, Duda A, Koshy B;
XX
XX WPI: 2002-106169/14.
XX P-PSDB: AAU11951.
XX
XX Isolated human pyridoxal (pyridoxine, vitamin B6) kinase polynts,
XX useful for therapeutic purposes, for studying the expression and
XX function of the polynt, and for expressing pyridoxal protein -
XX
XX Disclosure: Fig 1; 135pp; English.
```

```
CC The invention describes an isolated human pyridoxal (pyridoxine, vitamin
CC B6) kinase, (PDXK) polynucleotide. The polynucleotide is useful in
CC studying the expression and function of PDXK, and in expressing PDXK
CC protein for use in screening for candidate drugs to treat PDXK related
CC diseases and for therapeutic purposes. A transgenic animal is useful for
CC studying expression of the PDXK isogenes in vivo, for in vivo screening
CC and testing of drugs targeted against PDXK protein, and for testing the
CC efficacy of therapeutic agents and compounds for autoimmune polyglanular
CC disease type 1. The polypeptide is useful for studying the effect of the
CC variation on the biological activity of PDXK and the binding affinity of
CC candidate drugs targeting PDXK for the treatment of autoimmune
CC polyglanular disease type 1. Genotyping and haplotyping is useful for
CC improving the efficacy and reliability of several steps in the discovery
CC and development of drugs for treating diseases associated with PDXK
CC activity, e.g., autoimmune polyglanular disease type 1, to validate PDXK
CC as a candidate agent for treating a specific condition or disease
CC predicted to be associated with PDXK activity, and in the design of
CC clinical trials of candidate drugs. This sequence (located on chromosome
CC 21q22.3) encodes human pyridoxal (pyridoxine, vitamin B6) kinase (PDXK)
CC and forms the reference sequence on which the isoforms AAU11952 and
CC AAU11953 are based, described in the method of the invention.
```

Sequence 44861 BP; 9041 A; 12297 C; 13398 G; 10125 T; 0 other;

Query Match 9.4%; Score 47.2; DB 24; Length 44861;

Best Local Similarity 50.0%; Pred. No. 0.4; Mismatches 118; Indels 0; Gaps 0;

Matches 118; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
OY 236 CTGGGAGTGTGCGGTGCGCCGAGCGCCGCGGCGCGAGCAAGAGGAGCGCGG 295
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26041 CGCGGGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTCG 26100

OY 296 CCGCGAGCGGGGCCCGGAGCTTGCCTGCTCCTGCTGCGCCCGAGCGGGTTCGTCGCG 355
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26101 CGCGGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTCGCG 26160

OY 356 TTAGCGGACGAGCGCGCGCGGATGAGGCGGTAGCCCGGTGCCCCCTCGCGCGCAAG 415
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26161 GGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTCGCGGGG 26220

OY 416 CGCGCTGCGGCTGCGGCGCGGAGCTGCGCTGCTGCTGCGCGAGACAGCG 471
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26221 GGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTGCGGGGGGGCTCGAGGC 26276
```

```
RESULT 12
AAV62176/c
ID AAV62176 standard; DNA; 117213 BP.
XX
XX AAV62176;
XX
XX 13-JAN-1999 (first entry)
XX
```

```
XX HSV-2 strain SB5 Contig ID 15 DNA sequence.
XX
XX HSV-2 strain SB5; immunological response induction; therapy;
XX antiviral identification; viral protein inhibitor; ss.
XX
XX Herpes simplex virus type 2.
XX
XX Key Location/Qualifiers
XX CDS 755..1297
XX     /tag= a
XX     /product= "ORF#1 protein"
XX     /note= "encoded protein shown in AAW72170"
XX     1170..2174
XX     /tag= b
XX     /product= "ORF#2 protein"
XX     /note= "encoded protein shown in AAW72171"
XX     2229..2930
XX     /tag= c
XX     /product= "ORF#3 protein"
XX     /note= "encoded protein shown in AAW72172"
XX     complement (3130..3735)
XX     /tag= d
XX     /product= "ORF#4 protein"
XX     /note= "encoded protein shown in AAW72173"
XX     complement (3802..6447)
XX     /tag= e
XX     /product= "ORF#5 protein"
XX     /note= "encoded protein shown in AAW72174"
XX     6017..8482
XX     /tag= f
XX     /product= "ORF#6g protein"
XX     /note= "encoded protein shown in AAW72250"
XX     6026..8482
XX     /tag= g
XX     /product= "ORF#6f protein"
XX     /note= "encoded protein shown in AAW72249"
XX     6065..8482
XX     /tag= h
XX     /product= "ORF#6e protein"
XX     /note= "encoded protein shown in AAW72248"
XX     6167..8482
XX     /tag= i
XX     /product= "ORF#6d protein"
XX     /note= "encoded protein shown in AAW72247"
XX     6296..8482
XX     /tag= j
XX     /product= "ORF#6c protein"
XX     /note= "encoded protein shown in AAW72246"
XX     6326..8482
XX     /tag= k
XX     /product= "ORF#6b protein"
XX     /note= "encoded protein shown in AAW72245"
XX     6446..8482
XX     /tag= l
XX     /product= "ORF#6a protein"
XX     /transl_except= (pos: 7400..7402, aa: Ala-Ala)
XX     /transl_except= (pos: 7481..7486, aa: Ile)
XX     /note= "encoded protein shown in AAW72244"
XX     8457..9347
XX     /tag= m
XX     /product= "ORF#7 protein"
XX     /note= "encoded protein shown in AAW72175"
XX     complement (9604..11855)
XX     /tag= n
XX     /transl_except= (pos: 11635..11636, aa: Ala)
XX     /product= "ORF#8 protein"
XX     /note= "encoded protein shown in AAW72176"
XX     complement (11905..14508)
XX     /tag= o
XX     /product= "ORF#9b protein"
XX     /note= "encoded protein shown in AAW72222"
XX     complement (11905..14520)
XX
XX CDS
```

```
FT     /tag= p
FT     /product= "ORF#9a protein"
FT     /note= "encoded protein shown in AAW72223"
FT     14399..15802
FT     /tag= q
FT     /product= "ORF#10 protein"
FT     /note= "encoded protein shown in AAW72177"
FT     complement (15996..16286)
FT     /tag= r
FT     /product= "ORF#11 protein"
FT     /note= "encoded protein shown in AAW72178"
FT     complement (16202..18064)
FT     /tag= s
FT     /product= "ORF#12 protein"
FT     /note= "encoded protein shown in AAW72179"
FT     complement (18105..19661)
FT     /tag= t
FT     /product= "ORF#13 protein"
FT     /note= "encoded protein shown in AAW72180"
FT     complement (19415..20074)
FT     /tag= u
FT     /product= "ORF#14 protein"
FT     /note= "encoded protein shown in AAW72181"
FT     20155..21453
FT     /tag= v
FT     /product= "ORF#15 protein"
FT     /note= "encoded protein shown in AAW72182"
FT     complement (21326..22291)
FT     /tag= w
FT     /product= "ORF#16 protein"
FT     /note= "encoded protein shown in AAW72183"
FT     complement (22546..24654)
FT     /tag= x
FT     /product= "ORF#17 protein"
FT     /note= "encoded protein shown in AAW72184"
FT     24684..25955
FT     /tag= y
FT     /product= "ORF#18 protein"
FT     /note= "encoded protein shown in AAW72185"
FT     complement (26295..27251)
FT     /tag= z
FT     /product= "ORF#19 protein"
FT     /note= "encoded protein shown in AAW72186"
FT     complement (27630..31754)
FT     /tag= aa
FT     /product= "ORF#20b protein"
FT     /note= "encoded protein shown in AAW72225"
FT     complement (27630..31784)
FT     /tag= ab
FT     /product= "ORF#20a protein"
FT     /note= "encoded protein shown in AAW72224"
FT     complement (32067..32735)
FT     /tag= ac
FT     /product= "ORF#21 protein"
FT     /note= "encoded protein shown in AAW72187"
FT     33140..34984
FT     /tag= ad
FT     /product= "ORF#22a protein"
FT     /note= "encoded protein shown in AAW72226"
FT     33386..34984
FT     /tag= ae
FT     /product= "ORF#22b protein"
FT     /note= "encoded protein shown in AAW72227"
FT     complement (35205..37721)
FT     /tag= af
FT     /product= "ORF#23 protein"
FT     /note= "encoded protein shown in AAW72188"
FT     complement (38058..39188)
FT     /tag= ag
FT     /product= "ORF#24 protein"
FT     /note= "encoded protein shown in AAW72189"
FT     39090..39935
FT     /tag= ah
```

```
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in AAW72190"
FT CDS
FT /tag= a1
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in AAW72191"
FT CDS
FT /tag= a1
FT /transl_except= (pos: 44063..44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in AAW72192"
FT CDS
FT /tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in AAW72193"
FT CDS
FT /tag= a1
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in AAW72194"
FT CDS
FT /tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in AAW72195"
FT CDS
FT /tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in AAW72196"
FT CDS
FT /tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in AAW72197"
FT CDS
FT /tag= ap
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in AAW72198"
FT CDS
FT /tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in AAW72199"
FT CDS
FT /tag= ar
FT /product= "ORF#35 protein"
FT /note= "encoded protein shown in AAW72200"
FT CDS
FT /tag= as
FT /product= "ORF#36 protein"
FT /note= "encoded protein shown in AAW72201"
FT CDS
FT /tag= at
```

Query Match 9.4%; Score 46.8; DB 19; Length 117213;
Best Local Similarity 54.2%; Pred. No. 0.5;
Matches 116; Conservative 0; Mismatches 97; Indels 1; Gaps 1;

```
QY 229 TGTTCGCTTCGAGATGTCGCGGATGCCCGAGCGCGCGGAGCAAGGAGCG 288
DB 109471 TCCCGGCGGGGCGGTGGCGGGGCGGCGGCGGGAAGGAAAGAGCCCG 109412
QY 289 GAGCGGCGGAGCGGCGGCGGAGCTTCCTCCCTCGCTCGCCCGAGCGGGTTC 348
DB 109411 GAAGC-GCGGAGAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 109353
QY 349 GCTCGGCTAGAGCGAGCGCGCGCGATGAAGCGGTGAAGCCCGCGCCCTCGGGC 408
DB 109352 CGGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 109293
QY 409 CGCAAGCGCGCTTCGCGGCTCGCGCGCGGAGC 442
DB 109292 CGGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 109259
```

RESULT 13
AAF30757/c

```
ID AAF30757 standard; DNA; 47981 BP.
XX
XX AAF30757;
AC
XX 21-JUN-2001 (first entry)
DT
XX Micromonospora megalomicea megalomicin biosynthetic gene cluster.
DE
XX Megalomicin; meg gene; polyketide synthase; antibiotic;
XX motillide; antiparasitic; ds.
XX
XX Micromonospora megalomicea subsp. nigra.
OS
FH Key
FT CDS
FT Location/Qualifiers
FT complement (1..144)
FT /tag= a
FT /partial
FT /gene= "megBVI(megI)"
FT /product= "rnp-4-keto-6-deoxyglucose-
FT 2,3-dehydratase"
FT /note= "encodes AAB82201"
FT CDS
FT 928..2061
FT /tag= b
FT /gene= "megDVI"
FT /product= "rnp-4-keto-6-deoxyhexose 3,4-isomerase"
FT /note= "encodes AAB82202"
FT CDS
FT 2072..3382
FT /tag= c
FT /gene= "megDI"
FT /product= "rnp-megosamine glycosyltransferase"
FT /note= "eryCIII homologue; encodes AAB802203"
FT CDS
FT 3462..4634
FT /tag= d
FT /gene= "megI"
FT /product= "mycarose O-acyltransferase"
FT /note= "encodes AAB82204"
FT CDS
FT 4651..5775
FT /tag= e
FT /gene= "megDII"
FT /product= "rnp-3-keto-6-deoxyhexose
FT 3-aminoctransaminase"
FT 5822..6595
FT /tag= f
FT /gene= "megDIII"
FT /product= "daunosaminyl-N,N-dimethyltransferase"
FT /note= "eryCVI homologue; encodes AAB82206"
FT CDS
FT 6592..7197
FT /tag= g
FT /gene= "megDIV"
FT /product= "rnp-4-keto-6-deoxyhexose 3,5-epimerase"
FT /note= "eryBVIII, dnmU homologue; encodes AAB82207"
FT CDS
FT 7220..8206
FT /tag= h
FT /gene= "megDV"
FT /product= "rnp-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "eryYIV, dnmY homologue; encodes AAB82208"
FT CDS
FT complement (8228..9220)
FT /tag= i
FT /gene= "megBII-1(megDVI)"
FT /product= "rnp-4-keto-6-deoxyhexose 2,3-reductase"
FT /note= "encodes AAB82209"
FT CDS
FT complement (9226..10479)
FT /tag= j
FT /gene= "megBV"
FT /product= "rnp-mycarose glycosyltransferase"
FT /note= "encodes AAB82210"
FT CDS
FT complement (10483..11424)
FT /tag= k
FT /gene= "megBIV"
FT /product= "rnp-4-keto-6-deoxyhexose 4-ketoreductase"
FT /note= "encodes AAB82211"
FT CDS
FT 12181..22821
```

FT	/gene= "mega"
FT	/tag= l
FT	/product= "megalomycin 6-deoxyerithronolide B synthase 1"
FT	/note= "polyketide synthase; encodes AAB82213"
FT	misc_feature
FT	12505..13470
FT	/*tag= m
FT	/gene= "mega"
FT	/function= "At-L"
FT	13576..13791
FT	/*tag= n
FT	/gene= "mega"
FT	/function= "ACP-L"
FT	13849..15126
FT	/*tag= o
FT	/gene= "mega"
FT	/function= "KS1"
FT	15427..16476
FT	/*tag= p
FT	/gene= "mega"
FT	/function= "AT1"
FT	17155..17694
FT	/*tag= q
FT	/gene= "mega"
FT	/function= "KR1"
FT	17947..18207
FT	/*tag= r
FT	/gene= "mega"
FT	/function= "ACP1"
FT	18268..19548
FT	/*tag= s
FT	/gene= "mega"
FT	/function= "KS2"
FT	19876..20910
FT	/*tag= t
FT	/gene= "mega"
FT	/function= "AT2"
FT	21517..22053
FT	/*tag= u
FT	/gene= "mega"
FT	/function= "KR2"
FT	22318..22575
FT	/*tag= v
FT	/gene= "mega"
FT	/function= "ACP2"
FT	22867..33555
FT	/*tag= w
FT	/gene= "megAI1"
FT	/product= "megalomycin 6-deoxyerythronolide B synthase 2"
FT	/note= "polyketide synthase, encodes AAB82213"
FT	22957..24237
FT	/*tag= x
FT	/gene= "megAI1"
FT	/function= "KS3"
FT	24544..25581
FT	/*tag= y
FT	/gene= "megAI1"
FT	/function= "AT3"
FT	26230..26733
FT	/*tag= z
FT	/gene= "megAI1"
FT	/function= "KR3 (inactive)"
FT	26998..27258
FT	/*tag= aa
FT	/gene= "megAI1"
FT	/function= "ACP3"
FT	27393..28590
FT	/*tag= ab
FT	/gene= "megAI1"
FT	/function= "KS4"
FT	28897..29931
FT	/*tag= ac
FT	misc_feature

FT	/gene= "megAII"
FT	/function= "AT4"
FT	29953..30477
FT	/*tag= ad
FT	/gene= "megAII"
FT	/function= "DH4"
FT	31396..32244
FT	/*tag= ae
FT	/gene= "megAII"
FT	/function= "ER4"
FT	32257..32799
FT	/*tag= af
FT	/gene= "megAII"
FT	/function= "KR4"
FT	33052..33312
FT	/*tag= ag
FT	/gene= "megAII"
FT	/function= "ACP4"
FT	33666..43271
FT	/*tag= ah
FT	/gene= "megAII"
FT	/product= "megalomicin 6-deoxyerythronolide B synthase 3"
FT	/note= "polyketide synthase; encodes AAB82214"
FT	22957..24237
FT	/*tag= ai
FT	/gene= "megAII"
FT	/function= "KS5"
FT	24544..25581
FT	/*tag= aj
FT	/gene= "megAII"
FT	/function= "AW5"
FT	26230..26733
FT	/*tag= ak
FT	/gene= "megAII"
FT	/function= "KR5"
FT	26998..27258
FT	/*tag= al
FT	/gene= "megAII"
FT	/function= "ACPS"
FT	27393..28590
FT	/*tag= am
FT	/gene= "megAII"
FT	/function= "KS6"
FT	28897..29931
FT	/*tag= an
FT	/gene= "megAII"
FT	/function= "AT6"
FT	29953..30477
FT	/*tag= ao
FT	/gene= "megAII"
FT	/function= "KR6"
FT	31396..32244
FT	/*tag= ap
FT	misc_feature

Query Match	9.1%	Score 45.6	DB 22	Length 47981
Best Local Similarity	51.5%	Pred. 0.87	Mismatches 99	Indels 0
Matches 105	Conservative	0	Mismatches 99	Indels 0
QY 238	CGGAGGTGCGGGTCCCGGAGCCGCGCGCGCGGAGCGAAAGGAGCGGCGCC	297		
Db 17600	CCGCGGAGCCCGCGCGCGCGCGCAACCGCGGGGAGGAGGAGCAACGACCGTGG	17541		
QY 298	GGGAGCGGGGGCCGAGCGTTGGCTCTCTCCCTCGGTCGCGCCGAGCGGGTTCGCTGCCTA	357		
Db 17540	AGCTCGCGCTCCCGGGTCAAGCTCTGTCAGAGTTGGGGGCAACGAGCACTTCGCGCGTGG	17481		
QY 358	GAGCGCAGGGCGCGCGCGATGAAGCGCGTGAAGCCCGCGTTCGCGCCCGCAAGCGC	417		
Db 17480	GGCCCTTCGATCCGGTCAACCGGTAGGGTCTTCGACAGGTCCCTCTTCAGATCGCGGCG	17421		
QY 418	CCGTGCGGCTCGCGCGCGGAG	441		

OY 476 GCCTGGGTGGCTCC 489
|| ||| | |||
Db 408 GCGAGGGGAGTCC 421

Search completed: October 10, 2002, 14:54:20
Job time : 287 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 18.7333 Seconds
(without alignments)
6556.054 Million cell updates/sec

Title: US-09-489-101A-5_COPY_1_500

Perfect score: 500
Sequence: 1 atgctcctgagcgcgggtcc.....ccggcctgcagagcagca 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgnt2_6/prodata/2/ina/5A_COMB.seq.*
- 2: /cgnt2_6/prodata/2/ina/5B_COMB.seq.*
- 3: /cgnt2_6/prodata/2/ina/5A_COMB.seq.*
- 4: /cgnt2_6/prodata/2/ina/6B_COMB.seq.*
- 5: /cgnt2_6/prodata/2/ina/PCBUS_COMB.seq.*
- 6: /cgnt2_6/prodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113.2	22.6	3138	US-09-234-332-5	Sequence 5, Appl1
2	56	11.2	4403765	US-09-103-840A-2	Sequence 2, Appl1
3	55.4	11.1	44377	US-08-804-227C-7	Sequence 7, Appl1
4	55.4	11.1	44377	US-08-804-198-1	Sequence 1, Appl1
5	51.4	10.3	1910	PCT-US92-05532-1	Sequence 1, Appl1
6	51.4	10.3	2261	US-08-272-882D-1	Sequence 1, Appl1
7	51	10.2	2370	US-08-104-072B-7	Sequence 7, Appl1
8	51	10.2	2370	US-08-351-413-8	Sequence 8, Appl1
9	51	10.2	2370	US-09-025-583-8	Sequence 8, Appl1
10	51	10.2	5173	US-08-242-677-1	Sequence 1, Appl1
11	50.6	10.1	23673	US-09-773-816-1	Sequence 1, Appl1
12	50.4	10.1	28804	US-08-592-874-1	Sequence 1, Appl1
13	50.4	10.1	28804	US-09-096-942-2	Sequence 2, Appl1
14	50.4	10.1	28804	US-09-096-867-2	Sequence 2, Appl1
15	49.8	10.0	423	US-09-144-085-5	Sequence 5, Appl1
16	49.8	10.0	8438	US-07-945-283-1	Sequence 1, Appl1
17	49.8	10.0	13842	US-09-105-537-30	Sequence 30, Appl1
18	49.8	10.0	36778	US-09-105-537-5	Sequence 5, Appl1
19	49.8	10.0	38506	US-09-320-878-19	Sequence 19, Appl1
20	49	9.8	4257	US-08-690-473-1	Sequence 1, Appl1
21	49	9.8	4257	US-09-259-821A-1	Sequence 1, Appl1
22	49	9.8	4257	US-08-843-659-1	Sequence 1, Appl1
23	49	9.8	12001	US-08-458-568A-11	Sequence 11, Appl1
24	49	9.8	13987	US-08-804-227C-13	Sequence 13, Appl1
25	49	9.8	44377	US-08-804-227C-7	Sequence 7, Appl1
26	49	9.8	44377	US-08-804-198-1	Sequence 1, Appl1
27	48.6	9.7	4848	US-08-955-957A-1	Sequence 1, Appl1

28	48.6	9.7	4848	US-08-955-957A-4	Sequence 4, Appl1
29	48.6	9.7	4848	US-08-955-957A-6	Sequence 6, Appl1
30	48.4	9.7	1578	US-08-681-129-1	Sequence 1, Appl1
31	48.2	9.6	2721	US-08-5215881-2	Sequence 1, Appl1
32	48	9.6	1203	US-09-086-010-1	Sequence 1, Appl1
33	48	9.6	3836	US-08-216-260-1	Sequence 1, Appl1
34	48	9.6	4257	US-08-690-473-1	Sequence 1, Appl1
35	48	9.6	4257	US-09-259-821A-1	Sequence 1, Appl1
36	48	9.6	4257	US-08-843-659-1	Sequence 1, Appl1
37	48	9.6	12001	US-08-458-568A-11	Sequence 11, Appl1
38	47.8	9.6	1656	US-09-385-028-14	Sequence 14, Appl1
39	47.8	9.6	15079	US-09-385-028-1	Sequence 1, Appl1
40	47.6	9.5	1153	US-09-372-448A-5	Sequence 5, Appl1
41	47.6	9.5	2639	US-07-952-817-8	Sequence 8, Appl1
42	47.6	9.5	2639	US-07-952-817-8	Sequence 8, Appl1
43	47.6	9.5	3181	US-09-135-021-1	Sequence 1, Appl1
44	47.6	9.5	3181	US-09-135-020-1	Sequence 1, Appl1
45	47.6	9.5	3181	US-09-135-010A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-234-332-5
; Sequence 5, Application US/09234332A
; Patent No. 6087168
; GENERAL INFORMATION:
; APPLICANT: Cedars-Sinai Medical Center
; APPLICANT: Michael F. Levesque, M.D.
; APPLICANT: Thomas Neuman, Ph.D.
; TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
; FILE REFERENCE: P07 41496
; CURRENT APPLICATION NUMBER: US/09/234,332A
; CURRENT FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3138
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: gene
; LOCATION: (0)...(0)
; OTHER INFORMATION: Z1c 1 Protein gene; Genbank Accession D76435
US-09-234-332-5

Query Match	22.6%	Score 113.2;	DB 3;	Length 3138;
Best Local Similarity	63.3%	Pred. No. 1.4e-13;		
Matches 219;	Conservative	0;	Mismatches 103;	Indels 24;
Gaps 2;				
QY	139	CGCCGCGCATGAGGCTTAACTCAACCCGGGCGCGAGAGCTGCCGGGCA	218	
DB	900	CGCCGCGCATGAGGCTTAACTCAACCCGGGCGCGAGAGCTGCCGGGCA	959	
QY	219	GAGCTCGGGCTTACAGTGTGAGGCGGCGCTTACCCGGGCTGCGTCCGC	278	
DB	960	CGAGACAGCTTACAGTGTGAGGCGGCGCTTACCCGGGCTGCGTCCGC	1001	
QY	279	TGCGGCGCGAGGCTTACAGTGTGAGGCGGCGCTTACCCGGGCTGCGTCCGC	338	
DB	1002	TGCGGCGCGAGGCTTACAGTGTGAGGCGGCGCTTACCCGGGCTGCGTCCGC	1061	
QY	339	CAACTTCACCCGAGCTTCTTCCGAGCGGCGGCTTCCGGGACTTTCGGCGCGG	398	
DB	1062	CAACTTCACCCGAGCTTCTTCCGAGCGGCGGCTTTCGGGACTTTCGGCGCGG	1121	
QY	399	CGCGGCGCGAGGCTTTCGGGCGGCGGCGGCGCTTTCGGGCGGCGGCA	452	
DB	1122	CGCGGCGCGAGGCTTTCGGGCGGCGGCGGCGGCTTTCGGGCGGCGGCA	1181	
QY	453	CTCGGAGCGCGAGGCGGCGGCGGCGGCGGCTTTCGGGCGGCGGCA	498	

[illegible]

```

1      RESULT 6
2      US-08-272-882D-1
3      : Sequence 1, Application US/08272882D
4      : Patent No. 5516685
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Lichenstein, Henri
9      : TITLE OF INVENTION: Isolation and Characterization of No. 5516685e1
10     :
11     : NUMBER OF SEQUENCES: 7
12     : TITLE OF INVENTION: Plectase from Streptomyces lividans
13     :
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: Amgen Inc.
16     : STREET: 1840 Dehavenland Drive
17     : CITY: Thousand Oaks
18     : STATE: California
19     : COUNTRY: USA
20     : ZIP: 91320-1789
21     :
22     : COMPUTER READABLE FORM:
23     : MEDIUM TYPE: Floppy disk
24     : COMPUTER: IBM PC compatible
25     : OPERATING SYSTEM: PC-DOS/MS-DOS
26     : SOFTWARE: PatentIn Release #1.0, Version #1.30
27     :
28     : CURRENT APPLICATION DATA:
29     : APPLICATION NUMBER: US/08/272,882D
30     : FILING DATE:
31     : CLASSIFICATION: 435
32     :
33     : ATTORNEY/AGENT INFORMATION:
34     : NAME: Oleski, Nancy
35     : REFERENCE/DOCKET NUMBER: A-199A
36     :
37     : INFORMATION FOR SEQ ID NO: 1:
38     : SEQUENCE CHARACTERISTICS:
39     : LENGTH: 2261 base pairs
40     : TYPE: nucleic acid
41     : STRANDEDNESS: single
42     : TOPOLOGY: linear
43     :
44     : MOLECULE TYPE: DNA (genomic)
45     :
46     : US-08-272-882D-1

```

	Query Match	Similarity	10.3%	Score 51.4	DB 1	Length 2261
	Best Local	Similarity	46.1%	Pred. No. 0.043		
	Matches 172	Conservative	0	Mismatches 201	Indels	0
QY	118	CTGAGCCGTGGGGCGGCGACAGAACGGCTTCTGTGATTCGCCCGCGCCGACATGGAGCC	177			
Db	1192	CTGAGCGGCGCGCTGCTCGGACACGGACCCCGCGGAGCGCGCGCCACAGAGCGG	1251			
QY	178	TTCAAGCTTCACCCGGGGCGCGCACAGCTGTCCCGGGCGACAGCTCGGCGTTACAGTCG	237			
Db	1252	GTCAGAGCGCGCCCACTCGCTTACCAAGCCCTTCGGCGGCTCGGGGAGAGACCGGCGCGC	1311			
QY	238	CAGGGCCCCGGGGCGCTACCCCGGGGCTCGGTGGGGCTCGCGGCGCGCAAGGCTCGGG	297			
Db	1312	AACCGCGGCTTTTGAGAGCGGCTGCTCAAGTCCGTGCGGACAGAGGCGCGCGCAACCGG	1371			

[illegible]

```

1 RESULT 7
2 US-08-104-072B-7
3 : Sequence 7, Application US/08104072B
4 : Patent No. 5639948
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Michiels, Frank
9 : APPLICANT: MorioKa, Simji
10 : APPLICANT: Scheitlinck, Trees
11 : APPLICANT: Komari, Toshiko
12 : TITLE OF INVENTION: Stamen-specific Promoters from Rice
13 : NUMBER OF SEQUENCES: 38
14 :
15 : CORRESPONDENCE ADDRESS:
16 : ADDRESSEE: Merchant & Gould
17 : STREET: 3100 NO. 5639948west Center
18 : CITY: Minneapolis
19 : STATE: MN
20 : COUNTRY: USA
21 : ZIP: 55402
22 :
23 : COMPUTER READABLE FORM:
24 : MEDIUM TYPE: Floppy disk
25 : COMPUTER: IBM PC compatible
26 : OPERATING SYSTEM: PC-DOS/MS-DOS
27 : SOFTWARE: Patentin Release #1.0, Version #1.25
28 :
29 : CURRENT APPLICATION DATA:
30 : APPLICATION NUMBER: US/08/104,072B
31 : FILING DATE: 05-AUG-1993
32 : CLASSIFICATION: 800
33 :
34 : PRIOR APPLICATION DATA:
35 : APPLICATION NUMBER: WO 9200272
36 : FILING DATE: 06-FEB-1992
37 :
38 : PRIOR APPLICATION DATA:
39 : APPLICATION NUMBER: EP 91403352.7
40 : FILING DATE: 10-DEC-1991
41 :
42 : PRIOR APPLICATION DATA:
43 : APPLICATION NUMBER: EP 91402590.3
44 : FILING DATE: 27-SEP-1991
45 :
46 : PRIOR APPLICATION DATA:
47 : APPLICATION NUMBER: EP 91400318.1
48 : FILING DATE: 08-FEB-1991
49 :
50 : ATTORNEY/AGENT INFORMATION:
51 : NAME: Kowalczyk, Katherine M.
52 : REGISTRATION NUMBER: 36,848
53 : REFERENCE/DOCKET NUMBER: 8076.93USWO
54 :
55 : TELECOMMUNICATION INFORMATION:
56 : TELEPHONE: 612-332-5300
57 : TELEFAX: 612-332-9081
58 :
59 : INFORMATION FOR SEQ ID NO: 7:
60 :
61 : SEQUENCE CHARACTERISTICS:
62 : LENGTH: 2370 base pairs
63 : TYPE: nucleic acid
64 : STRANDEDNESS: double
65 : TOPOLOGY: linear
66 :
67 : MOLECULE TYPE: DNA (genomic)
68 :
69 : ORIGINAL SOURCE:
70 : ORGANISM: Oryza sativa

```

```
FEATURE:
NAME/KEY: promoter
LOCATION: 1..1808 /function= "another specific PT42
OTHER INFORMATION: promoter"
FEATURE:
NAME/KEY: TATA_signal
LOCATION: 1748..1755
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1780
OTHER INFORMATION: /product= "transcription
OTHER INFORMATION: initiation"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1809
OTHER INFORMATION: /product= "ATG start translation
OTHER INFORMATION: T42"
US-08-104-072B-7
```

```
Query Match 10.2% Score 51; DB 1; Length 2370;
Best Local Similarity 48.7%; Pred. No. 0.051;
Matches 167; Conservative 0; Mismatches 175; Indels 1; Gaps 1;
```

```
QY 7 CTGGACGGGGGTCGAGTTCCTCCGCGGAGTGGGGGAGCTTCGCGCGCCACCATCAC 66
DB 1799 CTAGACACCATGACAGATCATGAGCTCTCAAGGGGGGCGCCCTCTCTCTCATCT 1858
QY 67 CACTC-CGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 125
DB 1859 CATGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1918
QY 126 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 185
DB 1919 GCTGATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1978
QY 186 CAACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 245
DB 1979 CTGCTGGGAGAGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2038
QY 246 GCGCGCTTACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 305
DB 2039 CGTCTGCAATGGGTCAAGAGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2098
QY 306 CGCGCACGTTGGCTCTGCTGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 348
DB 2099 CTCGGCCCTCCCGCGCGCTGCGGGCTCTCTCATCAGCTTCAAC 2141
```

```
RESULT 8
US-08-351-413-8
Sequence 8, Application US/08351413
Patent No. 5750867
GENERAL INFORMATION:
APPLICANT: Williams, Mark
TITLE OF INVENTION: Maintenance of male-sterile plants
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESS: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 2046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/351.413
FILING DATE:
```

```
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2370 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Oryza sativa
STRAIN: Akihikari
FEATURE:
NAME/KEY: -
LOCATION: 1..1808 /label= PT42
OTHER INFORMATION: /note= "sequence comprising another specific
OTHER INFORMATION: promoter PT42"
FEATURE:
NAME/KEY: -
LOCATION: 1748..1755
OTHER INFORMATION: /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: -
LOCATION: 1780
OTHER INFORMATION: /note= "transcription initiation
OTHER INFORMATION: site determined by primer extension"
FEATURE:
NAME/KEY: -
LOCATION: 1809
OTHER INFORMATION: /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T42 gene"
US-08-351-413-8
```

```
Query Match 10.2% Score 51; DB 1; Length 2370;
Best Local Similarity 48.7%; Pred. No. 0.051;
Matches 167; Conservative 0; Mismatches 175; Indels 1; Gaps 1;
```

```
QY 7 CTGGACGGGGGTCGAGTTCCTCCGCGGAGTGGGGGAGCTTCGCGCGCCACCATCAC 66
DB 1799 CTAGACACCATGACAGATCATGAGCTCTCAAGGGGGGCGCCCTCTCTCTCATCT 1858
QY 67 CACTC-CGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 125
DB 1859 CATGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1918
QY 126 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 185
DB 1919 GCTGATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1978
QY 186 CAACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 245
DB 1979 CTGCTGGGAGAGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2038
QY 246 GCGCGCTTACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 305
DB 2039 CGTCTGCAATGGGTCAAGAGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2098
```


RESULT 15

US-09-144-085-5

; Sequence 5, Application US/09144085

; Patent No. 6280999

; GENERAL INFORMATION:

; APPLICANT: Gustafsson, Claes

; APPLICANT: Beliaich, Mary C.

; APPLICANT: Ashley, Gary

; APPLICANT: Julien, Bryan

; APPLICANT: Ziermann, Rainer

; TITLE OF INVENTION: SORANGIUM POLYPEPTIDE SYNTHASES AND ENCODING DNA

; TITLE OF INVENTION: THERFOR

; FILE REFERENCE: 30062-20020.20

; CURRENT APPLICATION NUMBER: US/09/144,085

; CURRENT FILING DATE: 1998-08-31

; EARLIER APPLICATION NUMBER: 09/010,809

; EARLIER FILING DATE: 1998-01-22

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 423

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE: Description of Artificial Sequence: DNA fragment

; OTHER INFORMATION: corresponding to a KS domain of Sorangium

; OTHER INFORMATION: cellosporium SMP44 gene

US-09-144-085-5

Query Match 10.0%; Score 49.8; DB 4; Length 423;

Best Local Similarity 47.6%; Pred.No.0.09; Mismatches 162; Indels 0; Gaps 0;

Matches 147; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

```
QY 1 ATGCTCTGGAGCGGGTCCGACGATTCGCGCCATGCGGGTGGGACGCTTCGGCGCGCCAC 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 ATGTTCGTGGAGTTCAGCCGTCAGCGGCGCGCTGCGCCGCGACGCGCAGCAGAGGCTTC 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 CATCACCACTCCGCGCGCGCGCGCGCGCTGCGCGCGAGATGACGACCGTGAACGTG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 TCGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGAGAGGCGCGCGCTGCTGCTGAGAGCG 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 AGCTTGGCGCGCGCGCGCGCGCGCGCTGCTGATTCGCGCGCGCGCGCATGGAGCCTTC 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 CTCTCGAGACGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 AAGCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 AACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 GGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 ATCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 CACGCGCGCG 309
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 CACGCGACG 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: October 10, 2002, 17:52:13
Job time : 1157.73 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 695.733 Seconds
(without alignments)
9693.805 Million cell updates/sec

Title: US-09-489-101a-5_COPY_1_500

Perfect score: 500

Sequence: 1 atgctcctgacgcggctcc.....ccggcctgcacagacagca 500

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	269	53.8	642	9	BB655427
2	224.2	44.8	278	9	BB655427
3	190.2	38.0	503	9	AT511335
4	115.7	31.4	496	9	AT559298
5	113.2	22.6	487	10	BE097311
6	108.4	21.7	441	9	BE097311
7	90	18.0	455	10	BE097311
8	86.8	17.4	935	12	BE097311
9	83.6	16.7	932	12	BE097311
10	80.2	16.4	925	12	BE097311
11	80.4	15.7	925	12	BE097311
12	78.4	15.4	821	12	BE097311
13	77.2	14.8	932	12	BE097311
14	73.6	14.7	976	12	BE097311
15	72.8	14.6	1008	12	BE097311
16	72.8	14.3	935	12	BE097311
17	71.6	14.3	935	12	BE097311

C 18	71.6	14.3	1203	12	CNS01514	AL106054 Drosophila
C 19	71.2	14.2	957	12	AG077010	AG077010 Pan trogl
C 20	71	14.2	885	12	AG159162	AG159162 Pan trogl
C 21	70.4	14.1	693	12	AG166257	AG166257 Pan trogl
C 22	70.2	14.0	652	10	BJ057546	BJ057546 BJO57546
C 23	70.2	14.0	872	12	CNS02418	AL188712 Tetradon
C 24	70	14.0	810	12	AG060267	AG060267 Pan trogl
C 25	70	14.0	841	12	A2196416	A2196416 SP.1031 B
C 26	70	14.0	922	12	AG057280	AG057280 Pan trogl
C 27	70	14.0	1021	12	AG061741	AG061741 Pan trogl
C 28	69.8	13.9	894	12	AG075627	AG075627 Pan trogl
C 29	69.6	13.9	813	12	A2193936	A2193936 SP.1025-A
C 30	69.4	13.9	1049	12	AG185368	AG185368 Pan trogl
C 31	69.2	13.8	894	12	AG075627	AG075627 Pan trogl
C 32	69.2	13.8	1452	12	AG032979	AG032979 Pan trogl
C 33	69	13.8	1104	12	AG043473	AG043473 Pan trogl
C 34	69	13.8	1160	12	AG043473	AG043473 Pan trogl
C 35	68.6	13.7	921	12	A2211117	A2211117 SP.0159 B
C 36	68.6	13.7	957	12	AG077010	AG077010 Pan trogl
C 37	68.6	13.7	1023	12	AG060164	AG060164 Pan trogl
C 38	68.6	13.7	1375	9	AW727483	AW727483 GA_Fa001
C 39	68.4	13.7	783	12	AG060185	AG060185 Pan trogl
C 40	68.4	13.7	1084	12	AG058435	AG058435 Pan trogl
C 41	68	13.6	663	12	AG080425	AG080425 Pan trogl
C 42	67.8	13.6	796	9	AG172073	AG172073 AUI72073
C 43	67.8	13.6	685	10	BG786336	BG786336 SEAMC006
C 44	67.8	13.6	1131	12	AG042920	AG042920 Pan trogl
C 45	67.6	13.5	1165	12	AG030649	AG030649 Pan trogl

ALIGNMENTS

RESULT 1
BB655427
LOCUS
DEFINITION
CDNA clone D030057J08 5', mRNA sequence.
ACCESSION
BB655427
VERSION
BB655427.1
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus.
REFERENCE
1 (bases 1 to 642)
Atakashi, T., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, H., Sasaki, Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Atakashi, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physiological and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayashi, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagui, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuda, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamazaki, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, T., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES	Location/Qualifiers
source	1. .642

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D030057J08"
/clone_1ib="RIKEN full-length enriched, 9 days embryo"
/dev_stage="9 days embryo"
/lab_host="DH10B"
/notes="Site.1: SalI; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGACAGACAGCGCCCGCCACACGAGTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermoactivated reverse transcriptase and subsequently enriched for full-length by cap trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGATTCGAGTTATTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambda FLX I."

```

BASE COUNT
ORIGIN

Query Match	53.8%	Score 269;	DB 9;	Length 642;
Best Local Similarity	85.4%	Pred. No. 6,3e-30;		
Matches	299;	Conservative	0;	Mismatches 51; Indels 0; Gaps 0;
QY	1	ATGCTCTGGACGCGGGGTCCGACAGTTTCCCGGCATCGGGGTGGGCACTTTCGGCGCCAC	60	
Db	293	ATGCTTTTGGACGCGGGGCCGACAGTTCCGGGCATCGGGGTGGGCACTTTCGGCGCCAC	352	
QY	61	CATCACCACTCCGCCCGCGGCGCGCGCGCGCTGCGCGCGAGACGAGACCTGTGACATG	120	
Db	353	CACCACTCACTCGCGCCGCGGACAGCGCGCGCTCGGGGCGCCGAGACGAGACCCGAGATG	412	
QY	121	AGCGTCGCGCGCGCGGAGAAAGCGTTGCTGATTTCCGCGCGCGCGGCAATGGAGGCTTC	180	
Db	413	AGCGTCGCGCGGACTTAAACCGGCTTCCTGGACATCGGCGCGGCGGCAATGGGCGCTTC	472	
QY	181	AACTCAACCCGGGCGCGGACGACTGTCCCGGGCCAGAGCTTGCGCTTCACGTCCAG	240	
Db	473	AAAGTCAAACCCGGGCGACAGCAAGACTGTCTCTGGTCAGAGTTTGGGTTTCACGTGCCA	532	
QY	241	GGCCCGCGCGCTACACCCCGGCTCCGCTTCGCGCTCGCGTCGCGCGGAGGCTCGGGGCC	300	
Db	533	GGTCGGGGTCTTACCCCGGGTTTCGGCTGCTGACGCGGTGGGCGCGGCTTAAAGGCC	592	
QY	301	CAGCGCGGACGTTGGCTCTACTCTGGGCGGCCCTTCAACTTCACACCG	350	
Db	593	CAGCGCGACAGTGGCTCTATTTCGAGGCTACCTTTAATTCACACCG	642	

RESULT 2
AI511335/C

1
2
3

LOCUS	A1511335	278 bp	mRNA	linear	EST 03-JUL-1998
DEFINITION	UI-R-C3- <i>sf-a</i> -03-0-UI.s1		UI-R-C3	Rattus norvegicus	cDNA clone
ACCESSION	A1511335				
VERSION	A1511335.1	GI:4417034			
KEYWORDS	EST.				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE

RODRIGUES	Normalization and subtraction: two approaches to facilitate gene discovery
TITLE	
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB

Tel: 319 335 8250
Fax: 319 335 9565
Email: mscores@iuiiue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Sears lab Clone distribution: clones will be available through
Research Genes (www.resgen.com) This clone is also available
through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov).
IMAGE ID=1768598
Seq Primer: M13 Forward
POLYA-No.

FEATURES

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C3-sf-a-03-0-0-UI"
/clone_idb="UI-R-C3"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pUT3p-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; The UI-R-C3
library is a subtracted library of a series, ultimately
derived from a mixture of individually tagged normalized
libraries from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, and 6, 12 and 16 day
embryos, after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: UI-R-C3, UI-R-C2p, UI-R-C1
, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C3) was constructed as follows: PCR amplified
cDNA inserts from UI-R-C2p clones from which 3' ESTs had
been derived was used as a driver in a hybridization with
the UI-R-C3p library in the form of single-stranded
circles. The remaining single-stranded circles (subtracted
library) was purified by hydroxyapatite column
chromatography, converted to double-stranded circles and
electroporated into DH10B bacteria (Life Technologies) to
generate the UI-R-C3 library. This procedure has been
previously described (Bonaldi, Lennon and Soares, Genome
Research 6:791-806, 1996)."

```

BASE COUNTRY
ORIGIN

Query Match	44.8%;	Score 224.2;	DB 9;	Length 278;
Best Local Similarity	88.1%;	Pred. No. 1.7e-23;		
Matches 244;	Conservative	0;	Mismatches 33;	Indels 0;
				Gaps 0.

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
12	AG063084	821 bp DNA	linear	GSS_03-NOV-2001		
	AG063084	Par troglodytes DNA, clone: PTB-051K02.R, genomic survey sequence.				
	AG063084	GI:16614866				
	GSS: (genome survey sequence).					
	Par troglodytes male lymphoblast DNA, clone lib:PTB Chimpanzee Male					

ORGANISM	REFERENCE
Pan troglodytes	Ekaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eulheta; Primates; Catarrhini; Homnidae; Pan.
1 (sites)	
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	
BAC end sequences of library PTB	
2 (bases 1 to 821)	
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	
Direct Submission	
Submitted (02-VUG-2001)	Asao Fujiyama, The Institute of Physical

COMMENT
Submitted (12 Aug 2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbpgsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT
Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS

```

Sequencing: M3Rev
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI
Location/Qualifiers
1. 821
FEATURES
source

```

	/cell_type="lymphoblast"
	/clone_id="PTB Chimpanzee Male BAC Library"
BASE COUNT	19 a 404 c 352 g 8 t 38 others
ORIGIN	

Query Match	15.7%;	Score 78.4;	DB 12;	Length 821;
Best Local Similarity	50.1%;	Pred. No. 0.013;		
Matches 221;	Conservative	0;	Mismatches 218;	Indels 2;
				Gaps 2;

QY	Db
70	251
311	

RESULT	13
LOCUS	AG081191.c
DEFINITION	Pan troglodytes DNA, clone: PTB-077108.F, genomic survey sequence.
ACCESSION	AG081191
VERSION	AG081191.1
KEYWORDS	GI:16632993
SOURCE	GSS: GSS (genome survey sequence).
REFERENCE	Pan troglodytes male lymphoblast DNA, clone_1b:PTB Chimpanzee Male

ORGANISM	REFERENCE
pan troglodytes	1 (sites)
Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.	
	1 (sites)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
Totoki, Y., Watanabe, H. and Sakaki, Y.	
BAC end sequences of library PTB	
Unpublished	
2 (bases 1 to 1369)	
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
Totoki, Y., Watanabe, H. and Sakaki, Y.	
Direct Submission	
Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical	
and Chemical Sciences, National Institute of Advanced Industrial	

COMMENT
1-7-22 Shuhiro Chou, Tsutsumi-Ku, Yokohama, Kanagawa 220-0045, Japan
(E-mail: chinpbe@gsc.riken.go.jp, URL: <http://bgp.gsc.riken.go.jp/>,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library pBn. This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS

```

Sequencing: -ZIM13
LIBRARY
  Vector      : pKS145
  R.Site 1    : SacI
  R.Site 2    : SacI.
Location/Qualifiers
  1. .1369
source

```

BASE COUNT	/clone_11b="PTB Chimpanzee Male BAC Library"				
ORIGIN	45 a	382 c	553 g	32 t	357 others

Query Match	15.48; Score 77.2; DB 12; Length 1369;
Best Local Similarity	47.28; Pred. No. 0.019;

	Matches	216,	Conservative	0;	Mismatches	241;	Indels	1;	Gaps	1,
Qy	12	C G C G G G T C G C A G T T T C C C G G C C - A T C G G G T G G G G C A G C T T T C G C G C C A C A T T A C C A C T	70							
Db	1295	C G C G C G C G C C C C C G C C C T T C C G G C C G C C T T G C G C N C G G C C G C C G C G C C C G C G C G C G C	1236							
Qy	71	C C G C C G A G A T G C A G S A O C G T G A A C T G A G C C T T G G C G G	130							
Db	1235	C G G G C G G G G G G G C C C A C G C C A G G C G C G C G C G C G C G C G C G C G C G C C C G G C C G G G C G C	1176							
Qy	131	C G G G C A G A A C G G C T T G T T G A T T C C C C G C C G C G C A C A T G G A G C C T T C A A G C T T A A C C	190							
Db	1175	C G C C C C G G G G C G C C G C G G C G G G C G G N C G C G C G G N C G C C G C G G G G G N C C G C G C C C C C G C	1116							
Qy	191	C G G G C G C G C A C A G A C T T C C C C G G G C C A G A G C T T C G C G T T A C G T C G C A G G G C C C G G G C	250							
Db	1115	G G C C C G G G C G C G C C C C C C C C C G G G C G C G G G G G G G C C G G C G G G A C G C G G C G G C	1056							
Qy	251	C C T A C C C C G G C T C C G T G C G G C T C C C C T G C G G C G C G A G C A C T T G G G G C C C A C C C C G C G C	310							
Db	1055	G C C C C C C N C G C A C N C C C C C C G C G C N C G N G N N C G C C G C C C C C G C G C C C G C C G C C	996							
Qy	311	A C G T T G G C T C T A C T T G G G C C G C C T T C A A C T C C A C C C G G A C T T C C T T T C C G A C G	370							
Db	995	N N C C G C C C G C G C G C G N N C C C G C C C C G C G C C C C C C C G C G C G C G C G C G C G C C	936							
Qy	371	C G C G C T T T C G G G G A C T T T G G G C G C G G G G C G G G A C A G C A C A G C A C T G T T G G G C G G G C G	430							
Db	935	G C G C G C G G G G C C G C C C C C G C C C C G C G C C G C G C G C G C G C G C G G G G C G G G C C G C G C G	876							
Qy	431	C G G G C G C C T C C A C C A C G C G C A C T C G A C G G C A G G C	468							
Db	875	C G G G C G G N G G G C G C G C G C C C C G C C C G C C C N N G G C	838							
RESULT 14	CNS00720/c									
LOCUS	CNS00720	932 bp	DNA	linear	GSS 03-JUN-1999					
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #									
LOCUS	BACR1B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.									
ACCESSION	AI066742									
VERSION	AI066742.1	GI:4945205								
KEYWORDS	GSS.									
SOURCE	fruit fly.									
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.									
REFERENCE	Genoscope. Direct Submision Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscp.cns.fr - Web : www.genoscope.cns.fr)									
AUTHORS	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).									
TITLE	The BDGP is constructing a physical map of the Drosophila melanogaster genome using three BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammossier at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.									
JOURNAL	Location/Qualifiers									
COMMENT	1. .932 /organism="Drosophila melanogaster"									
FEATURES	source									

BASE COUNT	155 a	202 c	241 g	91 t	243 others
ORIGIN	<pre> /db_xref="taxon:7227" /clone_11b="RPIC-98" /clone="BACR14B09" /notes="end : 737" </pre>				
Query Match	14.8%	Score 74.2	DB 12	Length 932	
Best Local Similarity	33.2%	Pred. No. 0.052			
Matches 141	Conservative 89	Mismatches 195	Indels 0	Gaps 0	
Qy	12	CGCGGGTCCGCAATTCCCGCCATCGGGGTGGGCACTTCGCGGCCACATACCACTC	71		
Db	835	CSGCGCCSCSSGCGSCGCCSCSSGCGSCSSCGCCSSCGCCSSCGCCSSCGCCSSCGC	776		
Qy	72	CGC	131		
Db	775	CGSCCGC	716		
Qy	132	GGCGCGAGACGGCTTCTTGTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA	191		
Db	715	CGGSCCGC	656		
Qy	192	GGC	251		
Db	655	GGC	596		
Qy	252	CTACCCCGCGCTCGCA	311		
Db	595	CMVAMAMASVSCCCSCCMASCCCGCGVSGCGSCSMSCCCCGCGASCVACGCM5AG	536		
Qy	312	CGTTGGCTCTACTCTGGC	371		
Db	535	CMGCGCGCMCCCGC	476		
Qy	372	GGC	431		
Db	475	CGVGMGCMGCMGRCVSNAGSGCAGMARMARAAACAGSGCAGMGCMGAGSAGCGVRG	416		
Qy	432	GGGCG 436			
Db	415	RAACG 411			
RESULT 15					
AG041117/c	976 bp DNA linear GSS 01-NOV-2001				
LOCUS	Pan troglodytes DNA, clone: PTB-018117.F, genomic survey sequence.				
DEFINITION	AG041117				
ACCESSION	AG041117.1 GI:16569842				
VERSION	GSS: GSS (genome survey sequence).				
KEYWORDS	Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male				
SOURCE	BAC library clone:PTB-018117.F.				
ORGANISM	Pan troglodytes				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.				
TITLE	1 (sites)				
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
REFERENCE	Totoki, Y., Watanabe, H. and Sakaki, Y.				
AUTHORS	BAC end sequences of library PTB				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 976)				
REFERENCE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
AUTHORS	Totoki, Y., Watanabe, H. and Sakaki, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical				
REFERENCE	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);				
AUTHORS	1-7-22 Suenho-chou, Tsukuba, Ibaraki, 305-8565, Japan				
TITLE	(E-mail:chimpesegsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/,				
JOURNAL	Tel:81-45-503-9111, Fax:81-45-503-9170)				
REFERENCE	Clones are derived from the chimpanzee BAC library PTB This BAC end				
AUTHORS	was generated during the R&D process and may have higher chance of				
TITLE	clone tracking errors.				

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .976

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-018L17.F"

/sex="male"

/cell_type="Lymphoblast"

/clone.lib="PTB Chimpanzee Male BAC Library"

BASE COUNT

59 a 287 c 495 g 51 t 84 others

ORIGIN

Query Match 14.7%; Score 73.6; DB 12; Length 976;
Best Local Similarity 48.2%; Pred. No. 0.064;
Matches 238; Conservative 0; Mismatches 245; Indels 11; Gaps 1;

```
QY 6 CTTGGAGCGCGGCTCCGAGTTCCCGGCCATCGGGGTGGGCACTTCGCGCCGACACATCA 65
    || || || || || || || || || || || || || || || || || || || || ||
Db 835 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 776
    || || || || || || || || || || || || || || || || || || || || ||
QY 66 CCACTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125
    || || || || || || || || || || || || || || || || || || || || ||
Db 775 NCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 716
    || || || || || || || || || || || || || || || || || || || || ||
QY 126 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 185
    || || || || || || || || || || || || || || || || || || || || ||
Db 715 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
    || || || || || || || || || || || || || || || || || || || || ||
QY 186 CAACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
    || || || || || || || || || || || || || || || || || || || || ||
Db 655 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 596
    || || || || || || || || || || || || || || || || || || || || ||
QY 246 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 305
    || || || || || || || || || || || || || || || || || || || || ||
Db 595 G-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
    || || || || || || || || || || || || || || || || || || || || ||
QY 306 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
    || || || || || || || || || || || || || || || || || || || || ||
Db 546 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 487
    || || || || || || || || || || || || || || || || || || || || ||
QY 366 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
    || || || || || || || || || || || || || || || || || || || || ||
Db 486 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 427
    || || || || || || || || || || || || || || || || || || || || ||
QY 426 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485
    || || || || || || || || || || || || || || || || || || || || ||
Db 426 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 367
    || || || || || || || || || || || || || || || || || || || || ||
QY 486 CTTGCGAGAGCAGC 499
    || || || || || ||
Db 366 CCGCGCGCGCGCGCGCG 353
```

Search completed: October 10, 2002, 20:27:22
Job time : 701.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 84.933 seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-5_COPY_1_500
Perfect score: 500
Sequence: 1 atgtcctgagcagcggtcc.....ccggcctgcagagcagca 500

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq.032802.*

1:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	1602	22	AAD11112 Human small cell 1
2	500	100.0	1602	24	AA561863 Lung small cell ca
3	113.2	22.6	3138	21	AAA62683 Human ZIC2 gene.
4	105.2	21.0	2947	24	ABI99239 Mouse ischaemic co
5	68.2	13.6	114955	20	AA53491 Human adenosine A1
6	64.2	12.8	114955	20	AA53491 Human adenosine A1
7	60	12.0	438	21	AA56076 Eucalyptus grandis
8	60	12.0	438	21	AA56076 Eucalyptus grandis
9	57	11.4	9968	22	AA136353 Human musculoskele

10	56.6	11.3	65140	22	AAD17184 Streptomyces nous
11	56.6	11.3	125401	22	AAD17186 Streptomyces nous
12	56.4	11.3	3957	22	AA409686 HSV-2 immediate ea
13	56.4	11.3	154746	24	AAD25519 Human herpesvirus
14	56.4	11.3	154746	24	AAD25519 Human herpesvirus
15	56	11.1	4403765	22	AA199683 Mycobacterium tube
16	55.6	11.1	3431	22	AA199683 Mycobacterium tube
17	55.4	11.1	37856	21	AA111992 S. cellulosum DNA
18	55.4	11.1	44377	18	AA178508 Platenolide syntha
19	55.4	11.1	44377	18	AA178508 Platenolide syntha
20	55	11.0	1811	23	AA591924 DNA encoding novel
21	54.6	10.9	1698	23	AA584408 DNA encoding novel
22	54.6	10.9	2049	21	AA576210 Nucleotide sequenc
23	53.4	10.7	58857	23	AA584711 Rabbit low density
24	53	10.6	2561	22	AA126500 Triticum sp. cyste
25	52.4	10.5	1791	22	AA126500 Triticum sp. cyste
26	52.4	10.5	1902	24	AA220003 Human HCN2 cDNA.
27	52.4	10.5	3372	21	AA148729 DNA encoding cent
28	51.8	10.4	3717	21	AA646560 Human ORF265
29	51.8	10.4	4566	21	AA646560 Human ORF265
30	51.4	10.3	1910	14	AA345459 Streptomyces livid
31	51.2	10.2	2061	12	AA186409 4-amino-4-deoxycho
32	51	10.2	2370	15	AA533880 Sequence compris
33	51	10.2	5173	18	AA533880 Human Immunodefici
34	50.6	10.2	3303	23	AA54127 Pseudomonas aerugi
35	50.6	10.1	1382	15	AA646500 Human derived adre
36	50.4	10.1	2561	22	AA126500 Sphingomonas S88 s
37	50.4	10.1	28804	17	AA126500 Sphingomonas S88 s
38	50.4	10.1	28804	18	AA126500 Sphingomonas S88 s
39	50.4	10.1	28804	20	AA126500 Sphingomonas S88 s
40	50.4	10.1	28804	20	AA126500 Sphingomonas S88 s
41	50.4	10.1	109519	22	AA508693 Chromosomal fragme
42	50.2	10.0	1053	22	AA444045 Streptomyces sp. C
43	50.2	10.0	1053	22	AA444045 Streptomyces sp. C
44	50.2	10.0	4563	21	AA178255 Nucleotide sequenc
45	50.2	10.0	4563	22	AA178255 Human PRO1487 (UNQ

ALIGNMENTS

RESULT 1	
ID	AAD11112 standard; DNA: 1602 BP.
XX	
XX	AAD11112:
XX	
DF	24-SEP-2001 (first entry)
XX	
DE	Human small cell lung cancer associated gene, ZIC2.
XX	
KW	Human; small cell lung cancer; therapy; hCAP; nucleic acid; NA;
KW	melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW	leiomyosarcoma; synovial sarcoma; cytotlastic; ZIC2; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	1..1602
FT	/*tag= a
FT	/product= "Human ZIC2 protein"
XX	
XX	
XX	W0200153349-A2.
XX	
PD	26-JUL-2001.
XX	
XX	19-JAN-2001; 2001WO-US02015.
XX	
XX	21-JAN-2000; 2000US-0489101.
XX	
XX	(LUDW-) LUDWIG INST CANCER RES.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.
PA	(CORR) CORNELL RES FOUND INC.

XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX WPI: 2001-457597/49.
DR P-PSDB: AAE05812.
XX Isolated polypeptide, used to treat or prognose a disorder
XX characterized by expression of a hCAP e.g. cancer, is encoded by an
XX isolated nucleic acid comprising an NA Group 3 or 4 molecule -
XX
XX Claim 57; Page 90-91; 152pp; English.
XX
XX The invention relates to nucleic acids and encoded polypeptides which
XX are cancer associated antigens expressed in patients afflicted with
XX small cell lung cancer. The molecules provided by the invention can be
XX used in the diagnosis, monitoring, research or treatment of conditions
XX characterized by the expression of one or more cancer associated
XX antigens. The polypeptide is used to treat a disorder characterized by
XX expression of a hCAP, and determine regression, progression or onset
XX of a condition characterized by expression of an abnormal amount of a
XX protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
XX are small and non-small cell lung cancer, melanoma, colon, breast, head
XX and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
XX The present sequence is a small cell lung cancer associated gene
XX designated as NY-SCC-3 encoding human ZIC2 protein.
XX
SQ Sequence 1602 BP; 275 A; 578 C; 536 G; 213 T; 0 other;
Query Match 100.0%; Score 500; DB 22; Length 1602;
Best Local Similarity 100.0%; Pred. No. 3e-75;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTCTGGAGCGGGGCTCCAGTTCGGGCGCATCGGGGTTGGGAGCTTCCGGCCAC 60
Db 1 ATGCTCTGGAGCGGGGCTCCAGTTCGGGCGCATCGGGGTTGGGAGCTTCCGGCCAC 60
QY 61 CATCACACTCCG 120
Db 61 CATCACACTCCG 120
QY 121 AGCTGGGGGGGGGCGAAGGGCTTGTGATTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 AGCTGGGGGGGGGCGAAGGGCTTGTGATTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 AAGCTCAACCGGGGCG 240
Db 181 AAGCTCAACCGGGGCG 240
QY 241 GGGCGGGGCGCTACCG 300
Db 241 GGGCGGGGCGCTACCG 300
QY 301 CAGCGCGCGCGAGTGGCTTACTCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 301 CAGCGCGCGCGAGTGGCTTACTCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 TTCCGACAGCGGGGCTCCGGGAGCTCCGGGCGCGGGGCGGGGCGCGGAGTGGTTC 420
Db 361 TTCCGACAGCGGGGCTCCGGGAGCTCCGGGCGCGGGGCGGGGCGCGGAGTGGTTC 420
QY 421 GGGCGGGGCGGGGCGCGCTGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 421 GGGCGGGGCGGGGCGCGCTGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 CCGGGCGCGCGAGAGCA 500
Db 481 CCGGGCGCGCGAGAGCA 500
RESULT 2
AAS61863
ID AAS61863 standard; cDNA; 1602 BP.
XX

AC AAS61863;
XX
XX 29-JAN-2002 (first entry)
XX
XX Lung small cell carcinoma antigen, cDNA #404.
XX
XX Human; cytosolic; antitumour; lung small cell cancer antigen;
XX tumour; lung cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200177168-A2.
XX
XX 18-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US11859.
XX
XX 11-APR-2000; 2000US-196780P.
XX 21-JUN-2000; 2000US-21361P.
XX 01-SEP-2000; 2000US-229763P.
XX 05-SEP-2000; 2000US-230629P.
XX 14-SEP-2000; 2000US-232655P.
XX 19-DEC-2000; 2000US-257037P.
XX 08-JAN-2001; 2001US-260796P.
XX
XX (CORI-) CORIYA CORP.
XX
XX Lodes MJ, Wang T, Mohamath R, Indrias CY;
XX WPI: 2002-010896/01.
XX P-PSDB; AAU69423.
XX
XX Lung tumour polynucleotide and polypeptides useful in therapy and
XX diagnosis of cancer especially lung cancer -
XX
XX Claim 1; Page 274; 295pp; English.
XX
XX The invention relates to novel isolated lung small cell cancer antigen
XX polynucleotides (I) and polypeptides (II) used in a method of detecting
XX cancer in a patient. The method is optionally performed by
XX utilizing oligonucleotides (III), where the biological sample
XX from the patient is contacted with (III), detecting the amount of
XX polynucleotide hybridised to (III) in the sample and comparing the
XX amount of polynucleotide to a predetermined cut-off value and thereby
XX determining cancer in a patient. (I), (II) or antigen-presenting cells
XX expressing (II) is useful for stimulating and/or expanding T cells
XX specific for a tumour protein. The method comprises contacting T cells
XX with one of the components under conditions to permit the stimulation
XX and/or expansion of the cells. A composition comprising (I) is useful for
XX stimulating an immune response in a patient and for inhibiting the
XX development of a cancer especially lung cancer in a patient. An
XX isolated T cell population is useful for removing tumour cells from the
XX biological sample and for inhibiting the development of cancer in a
XX patient. AAS61460-AAS61874 represent novel human lung small cell
XX cancer antigen coding sequences of the invention.
XX
SQ Sequence 1602 BP; 275 A; 578 C; 536 G; 213 T; 0 other;
Query Match 100.0%; Score 500; DB 24; Length 1602;
Best Local Similarity 100.0%; Pred. No. 3e-75;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTCTGGAGCGGGGCTCCAGTTCGGGCGCATCGGGGTTGGGAGCTTCCGGCCAC 60
Db 1 ATGCTCTGGAGCGGGGCTCCAGTTCGGGCGCATCGGGGTTGGGAGCTTCCGGCCAC 60
QY 61 CATCACACTCCG 120
Db 61 CATCACACTCCG 120
QY 121 AGCTGGGGGGGCGAAGGGCTTGTGATTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 AGCTGGGGGGGCGAAGGGCTTGTGATTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180

```

QY 181 AAGCTCAACCCGGGCGGCGAGCTGTCCCGGGCCAGAGCTCGGCTTACGTCGAG 240
Db 181 AAGCTCAACCCGGGCGGCGAGCTGTCCCGGGCCAGAGCTCGGCTTACGTCGAG 240
QY 241 GGGCCGGGCGCTTACCCGGCTCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 300
Db 241 GGGCCGGGCGCTTACCCGGCTCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 300
QY 301 CAGCGCGGCGAGCTGTGCTCTTACTGTGCGGCGGCTTCACTTCCAGCCGGGACTTCTG 360
Db 301 CAGCGCGGCGAGCTGTGCTCTTACTGTGCGGCGGCTTCACTTCCAGCCGGGACTTCTG 360
QY 361 TTCCGCAACCGGCGGCTTCCGCGGAGCTTCCGCGGCGGCGGCGGCGGCGGCTTTC 420
Db 361 TTCCGCAACCGGCGGCTTCCGCGGAGCTTCCGCGGCGGCGGCGGCGGCGGCTTTC 420
QY 421 GGGCGGGGCGGCGGCGGCTTCCGCGGAGCTTCCGCGGCGGCGGCGGCGGCTTTC 480
Db 421 GGGCGGGGCGGCGGCGGCTTCCGCGGAGCTTCCGCGGCGGCGGCGGCGGCTTTC 480
QY 481 CCGGCGCTGCGAGAGCAGCA 500
Db 481 CCGGCGCTGCGAGAGCAGCA 500

RESULT 3
AAA62683
ID AAA62683 standard; DNA: 3138 BP.
XX
AC AAA62683;
XX
DT 29-NOV-2000 (first entry)
XX
DE Human Zic1 gene.
XX
KW Human; Zic1; epidermal cell transdifferentiation; gene therapy;
KW cerebroprotective; neuroprotective; brain injury; spinal cord injury;
KW stroke; neurodegenerative disease; Parkinson's disease;
KW Huntington's disease; Alzheimer's disease; neuronal cell generation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 781..2124
FT /lag- a
FT /product= "Zic1"
XX
PN US6087168-A.
XX
PD 11-JUL-2000.
XX
PE 20-JAN-1999; 99US-0234332.
XX
PR 20-JAN-1999; 99US-0234332.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Levesque MF, Neuman T;
XX
DR WPI: 2000-498200/44.
XX
DR P-PSDB: AAB14349.
XX
PT Converting epidermal cells into neurons, useful for isolating nerve
PT growth factors or for gene therapy, comprises differentiating cells
PT and transfecting with vectors with a cDNA coding neurogenic
PT transcription factors
XX
PS Example 2; Column 23-26; 27pp; English.
XX
CC The present sequence is the human Zic1 gene from Genbank. It was
CC used to provide sequence information for the cloning of Zic1 cDNA,
CC which was used to transfect cultured epidermal cells. This was

```

```

CC part of a novel method for transdifferentiating an epidermal basal
CC cell into a cell having the morphological, physiological and/or
CC immunological features of a viable neuronal cell. The method is useful
CC for screening new drugs for treating a nervous system disorder, or for
CC isolating a novel nerve growth factor. The transdifferentiated cell is
CC useful in both cell and gene therapies aimed at alleviating various
CC neurological disorders. The cell or gene therapy approach involves the
CC use of autologous transplantation or grafting of the newly created
CC neuronal cells as treatment for brain or spinal cord injury, stroke and
CC neurodegenerative diseases (e.g. Parkinson's disease, Huntington's
CC disease or Alzheimer's disease).
XX
SQ Sequence 3138 BP; 701 A; 898 C; 840 G; 699 T; 0 other;
Query Match 22.6%; Score 113.2; DB 21; Length 3138;
Best Local Similarity 63.3%; Pred. No. 6,4e-11;
Matches 219; Conservative 0; Mismatches 103; Indels 24; Gaps 2;
QY 159 CGCGCGCAGATGAGAGCTTCAAGCTCAACCGGCGGCGAGAGCTGTCCCGGGCCA 218
Db 900 CGCGGAGGCGATGAGGCGCTTCAAGCTCAACCGGCGGCGAGAGCTGTCCCGGG 959
QY 219 GAGCTGCGGCTTCAAGCTGCGAGGCGGCGGCTTCAAGCTGCGGCTGCGGCTGCGG 278
Db 960 CAGAGACGCTTCAAGCTGCGAGGCGGCGGCTTCAAGCTGCGGCTGCGGCTGCGG 1001
QY 279 TGGCGCGGCGAGGCTTGGGCGGCGGCGGCGGCGGCTTCAAGCTGCGGCTGCGG 338
Db 1002 TGGCGCGGCGAGGCTTGGGCGGCGGCGGCGGCGGCTTCAAGCTGCGGCTGCGG 1061
QY 339 CAAGCTCAAGCGGAGCTTCTGTTCCGAGCGGCGGCTTCCGCGGAGCTTCCGCGG 398
Db 1062 CAAGCTCAAGCGGAGCTTCTGTTCCGAGCGGCGGCTTCCGCGGAGCTTCCGCGG 1121
QY 399 CGGCGGCGAGCAGCGGCTTGGGCGGCGGCGGCGGCGGCTTCAAGCTGCGGCTG 452
Db 1122 CAGCGGCGAGCAGCGGCTTGGGCGGCGGCGGCGGCGGCTTCAAGCTGCGGCTG 1181
QY 453 CTGCGAGCGGCGAGGCGGCTTCCGCGGCGGCGGCGGCGGCTTCAAGCTGCGGCTG 498
Db 1182 CAGCGGCGGCGGCGGCGGCTTCCGCGGCGGCGGCGGCGGCTTCAAGCTGCGGCTG 1227

RESULT 4
AB199239
ID AB199239 standard; cDNA: 2947 BP.
XX
AC AB199239;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO: 60.
XX
DE Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
PN W0200188188-A2.
XX
PD 22-NOV-2001.
XX
PE 18-MAY-2001; 2001WO-JP04192.
XX
PR 18-MAY-2000; 2000JP-0145977.
XX
PA (UYN1-) UNIV NITON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
DR WPI: 2002-034733/04.
DR P-PSDB: ABB57042.
XX

```

PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or
 PT by determining the expression profile of a gene group comprising these
 PT genes -

PS Claim 2: Page 191-194; 260pp; English.

CC The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (1) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (1). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 CC expression levels of particular genes (AB199202 to AB199912, encoding
 CC the protein sequences in AB199202 to AB199912) or by determining the
 CC expression profile of a gene group comprising these genes. The
 CC expression levels or expression profiles produced by these genes are
 CC used as an indicator when screening for ischaemic condition-improving
 CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914
 CC represent PCR primers for a mouse ischaemic condition related sequence,
 CC which are used in the exemplification of the present invention.

SO Sequence 2947 BP; 659 A; 818 C; 787 G; 683 T; 0 other;

Query Match 21.0%; Score 105.2; DB 24; Length 2947;
 Best Local Similarity 61.8%; Pred. No. 1.4e-09;
 Matches 214; Conservative 0; Mismatches 108; Indels 24; Gaps 2;

OY 159 CGCGGCGACATGGAGCTTCAAGCTCAACCGGGGCGGACAGAGCTGTCCCGGCGCA 218
 DB 650 CGCGGCGACATGGAGCTTCAAGCTCAACCGGGGCGGACAGAGCTGTCCCGGCGCG 709
 OY 219 GAGCTGTTCAGCTGCGAGGGCGCGGCTTACCGCGGCTCGGCGGCTGCGCGG 278
 DB 710 CCAAGACAGCTTCAAGCTTCCAGAGCTCGGCGCTACCGGCTACCGGCTACCGG 751
 OY 279 TCGCGGCGGAGCGCTGCGGCGCGGCGGCGGCTTCTCTCTCTCTCTCTCTCTCT 338
 DB 752 TCGCGGCGGAGCGCTGCGGCGCGGCGGCGGCTTCTCTCTCTCTCTCTCTCTCT 811
 OY 339 CAAGCTCAACCGGGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 398
 DB 812 CAATTTACCGGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 871
 OY 399 CGCGGCGGACAGCGGCTGTGCGGCGGCGGCGGCGCTTCTCTCTCTCTCTCTCT 452
 DB 872 TAGCGGCGGAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 931
 OY 453 CTGCGGCGGAGCGGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 498
 DB 932 TAGGAGCGGCGGCGGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 977

RESULT 5
 AAX53491
 ID AAX53491 standard; DNA; 114955 BP.

AC AAX53491;

DT 05-JUL-1999 (first entry)

XX Human adenosine A1 receptor antisense oligonucleotide fragment.
 XX
 XX Antisense oligonucleotide; multiple target; antisense treatment;
 KW implied respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.

XX
 OS Synthetic.
 XX
 PN WO9913886-A1.
 XX
 XX 25-MAR-1999.
 XX
 PF 17-SEP-1998; 98WO-US19419.
 XX
 PR 09-JUN-1998; 98US-0093972.
 XX
 PR 17-SEP-1997; 97US-0059160.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Myce JW;
 XX
 DR WPI: 1999-229400/19.
 XX
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 XX
 PS Disclosure; Page 37; 120pp; English.

CC The specification describes antisense oligonucleotides (AAX52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of mRNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of mRNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences AAX55272-74. These multiple target
 CC oligonucleotides (specifically AAX55180-271) can be used for the
 CC antisense treatment of diseases and conditions. Typical diseases and
 CC conditions are those associated with impaired respiration and
 CC inflammation, including lung diseases, pulmonary vasoconstriction,
 CC inflammation, allergic rhinitis, acute asthma, allergy, asthma, impeded
 CC respiration, respiratory distress syndrome, pain, cystic fibrosis,
 CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic
 CC obstructive pulmonary disease (COPD), and cancers such as leukemias,
 CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,
 CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,
 CC hepatic metastases, as well as all types of cancers which may metastasize
 CC or have metastasized to the lungs, including breast and prostate cancer.

XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;
 Query Match 13.6%; Score 68.2; DB 20; Length 114955;
 Best Local Similarity 36.8%; Pred. No. 0.0013;
 Matches 175; Conservative 32; Mismatches 267; Indels 1; Gaps 1;

OY 15 GGGTCGCGAGTTCGCGGCGGATCGGGGTGGGAGCTTCGCGGCGGACATACACATCCCG 74
 DB 105005 BGGCCBGGGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 105064
 OY 75 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 134
 DB 105065 CGSNNNDNNGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 105124
 OY 135 GCGAGACGGCTTCGTTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 194
 DB 105125 CGGGCGSNNNDNNGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 105184
 OY 195 CGCGGCGGAGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 254
 DB 105185 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105244
 OY 255 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 314
 DB 105245 GGGCGSNNNDNNGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 105304
 OY 315 TGGCTCTACTCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 373
 DB 105305 NNDNNGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105364


```

XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
PS Claim 1, Page 100; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 438 BP; 66 A; 196 C; 111 G; 65 T; 0 other;
Query Match 12.0%; Score 60; DB 21; Length 438;
Best Local Similarity 51.1%; Pred No. 0.059;
Matches 141; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
XX
XX 39 GGTGGGCAAGCTTCGGCGCCACCAATCACCCTCCGCGCGGCGGCGGCTGCCGC 98
DB 130 GGGCGGCGCCACCAAGGCGCCAGCACCCTCCGCGCGGCGGCGGCGGCGG 189
XX
XX 99 CGAGATGACGAGACCGCTGATGAGCTGGCGGCGGCGGAGACGAGCTTCGTTGATTCGC 158
DB 190 TGACGACACGCGGACACCTCGAGACCTCCGCGGCGAGACCTCTCCGCGTCTCCGG 249
XX
XX 159 CGCGCGGCGACATGAGAGCTTCAGAGTCAACCGGCGGCGGCGGAGCTGTCCCGGCGCA 218
DB 250 GAAGGTCCACGAGTGTGATGTCCACAGAGCTTCCACCGGCGGCGGCTGCGGCGG 309
XX
XX 219 GAGCTCGGCGTTACGTCGACGAGGCGCCGCGGCGCTACCGCGGCTCGCTGCGGCGC 278
DB 310 GCACAAGCGGCTGCTGATGTCCACAGAGCTTCCCGCGGCTCTCTCCGCGGCGC 369
XX
XX 279 TCGGCGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314
DB 370 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 405
XX
XX RESULT 8
XX AAC56714
XX ID AAC56714 standard; DNA; 438 BP.
XX
XX AAC56714;
XX
XX 25-JAN-2001 (first entry)
XX
XX Eucalyptus grandis transcription factor DNA sequence #585.
XX
XX DE Eucalyptus grandis transcription factor DNA sequence #585.
XX
XX KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
XX KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
XX type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX OS Eucalyptus grandis.
XX
XX PN WO200053724-A2.
XX
XX PD 14-SEP-2000.
XX
XX PF 09-MAR-2000; 2000WO-US06112.
XX
XX PR 11-MAR-1999; 99US-0266513.

```

```

PR 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shank MA, Glenn M;
XX WPI: 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
PS Claim 1, Page 498; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 438 BP; 65 A; 195 C; 111 G; 65 T; 2 other;
Query Match 12.0%; Score 60; DB 21; Length 438;
Best Local Similarity 51.1%; Pred No. 0.059;
Matches 141; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
XX
XX 39 GGTGGGCAAGCTTCGGCGCCACCAATCACCCTCCGCGCGGCGGCGGCGGCTGCCGC 98
DB 130 GGGCGGCGCCACCAAGGCGCCAGCACCCTCCGCGCGGCGGCGGCGGCGG 189
XX
XX 99 CGAGATGACGAGACCGCTGATGAGCTGGCGGCGGCGGAGACGAGCTTCGTTGATTCGC 158
DB 190 TGACGACACGCGGACACCTCGAGACCTCCGCGGCGAGACCTCTCCGCGTCTCCGG 249
XX
XX 159 CGCGCGGCGACATGAGAGCTTCAGAGTCAACCGGCGGCGGCGGAGCTGTCCCGGCGCA 218
DB 250 GAAGGTCCACGAGTGTGATGTCCACAGAGCTTCCCGCGGCGGCGGCTGCGGCGG 309
XX
XX 219 GAGCTCGGCGTTACGTCGACGAGGCGCCGCGGCGCTACCGCGGCTCGCTGCGGCGC 278
DB 310 GCACAAGCGGCTGCTGATGTCCACAGAGCTTCCCGCGGCTCTCTCCGCGGCGC 369
XX
XX 279 TCGGCGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314
DB 370 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 405
XX
XX RESULT 9
XX AAL36353/C
XX ID AAL36353 standard; DNA; 9968 BP.
XX
XX AAL36353;
XX
XX 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 2718.
XX
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 2718.
XX
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.

```

OS Homo sapiens.
 XX PN W0200155367-A1.
 XX PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01338.
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217486.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225274.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227189.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235634.
 PR 27-SEP-2000; 2000US-0235636.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237040.
 PR 02-OCT-2000; 2000US-0237049.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.

CC Involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
 XX

SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

Query Match 11.3%; Score 56.6; DB 22; Length 125401;
 Best Local Similarity 44.6%; Pred. No. 0.11;
 Matches 221; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

2 TGGTCTGAGACGGGTCGCACTCCGCGCATCAGGCTGGGAGCTTCGCGCCAC 61
 79248 TCGTGTCCGCGATGCGCCAGAGATCCCTGCGAGCGGCGCCGCGCGCTGTGCG 79307
 62 ATACACCACTCCGCGGCGGCGGCGGCTGCGCGAGATGAGACCTGAACG 121
 79308 GCAAGACCGCCCGAGAGACTCTCGCGGTACCGGCTGCGCGCGCGCGCG 79367
 122 GCGTGGCGGCGGCGAGAACGGCTTGATTCCGCGCGCGCGACATGGAGAGCTTCA 181
 79368 GCGTACAGGTCCTGCTGGCGCGCTCTGACGCGACCGCGCGCGCGCGACTGC 79427
 182 AGCTCAACCGGCGGCGGCGAGCTGTCGCCGCGCGAGAGCTCGGCGCTTACAGTGCAG 241
 79428 CCACCTTACCTTTCAGACACAGCGTTCTGCGCGACCGCGCGCGCGCGCGCG 79487
 242 GCGCGCGCGCTTACCGCGCTGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCG 301
 79488 TCACCGCGCGGCGGAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79547
 302 ACAGCGCGCGAGTGGCTCTTACTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 361
 79548 CCGAGCGGCGCGGCTTACTTCTTACAGCGCGCGCTCTCGCGAGCCCGCGCGCG 79607
 362 TCGCGACGCGCGGCTTCTCGGAGACTTGGCGCGCGCGCGCGCGCGCGCGCGCG 421
 79608 CCGACACGCGGCTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79667
 422 GCGCGCGCGGCGGCGGCTTCAACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 481
 79668 TCGCGCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79727
 482 CGGCGCGCGCGAGG 496
 79728 TGGTGTGCGCGAGC 79742

RESULT 12
 AAA09686
 ID AAA09686 standard; DNA; 3957 BP.
 XX
 AC AAA09686;
 XX
 DT 31-JAN-2001 (first entry)
 XX
 DE HSV-2 immediate early protein ICP4 DNA sequence.
 XX
 KW Herpes-simplex-virus type 2; HSV-2; infected cell protein 4; ICP4;
 KW vaccine; infection; ds.
 XX
 OS Herpes simplex virus type 2.
 XX
 PN MO9516779-A1.
 XX
 PD 22-JUN-1995.
 XX
 PF 13-DEC-1994; 94MO-BP04138.
 XX
 FN 14-DEC-1993; 93GB-0025496.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Pala P, Gheysen DR, Slaoui MM, Koutoukos MC;

XX WPI: 2001-024142/03.
 DR P-PSDB; AAB26874.
 XX

PT Immediate early herpes-simplex-virus type 2 (HSV-2) ICP4 protein is
 PT used in vaccines for therapeutically or prophylactically treating HSV
 PT infections -
 XX
 PS Claim 5; Page 16; 28pp; English.
 XX

CC This invention relates to an immediate early herpes-simplex-virus type 2
 CC (HSV-2) infected cell protein 4 (ICP4) recognised by human cytotoxic T
 CC cells. HSV-2 ICP4 protein is recognized by cytotoxic T-lymphocyte (CTL)
 CC cells in humans and is used in vaccines for therapeutically or
 CC prophylactically treating HSV infections. Pharmaceutical compositions of
 CC HSV-2 ICP4 protein may be used to treat patients suffering from HSV
 CC infections, to prevent or decrease recurrent herpes disease, frequency,
 CC severity and duration of episodes. The present sequence represents HSV-2
 CC DNA encoding ICP4.
 XX

SQ Sequence 3957 BP; 368 A; 1656 C; 1568 G; 365 T; 0 other;

Query Match 11.3%; Score 56.4; DB 22; Length 3957;
 Best Local Similarity 45.6%; Pred. No. 0.18; Mismatches 236; Indels 0; Gaps 0;
 Matches 198; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

54 GCGCACCATCACACTCCGCGCGGCGGCGGCGGCTGCGCGCGCGAGATGACGACG 113
 2127 GCGCTCGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2186
 114 TGAACGTGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173
 2187 GCGCGTGAAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2246
 174 AGCTTCAAGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233
 2247 GACTTCCG 2306
 234 GTGCGAGGCG 2393
 2307 GCTGCG 2366
 294 CGGCG 353
 2367 GCG 2426
 354 CTTCCTGTTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 413
 2427 CCGGAAGAGAGCG 2486
 414 GCTGTTGCGGCG 473
 2487 GCG 2546
 474 CCGTTCG 487
 2547 GCGCGCGCGAGGCGC 2560

RESULT 13
 AAD25519
 ID AAD25519 standard; DNA; 154746 BP.
 XX
 AC AAD25519;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human herpesvirus 2 complete DNA genome.
 XX
 KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
 KW antibacterial; fungicide; protozoicide; antiparasitic; antiinflammatory;
 KW antitubercular; rheumatoid arthritis; neuroprotective; multiple sclerosis;
 KW immune response; vasotrophic; vaccine; gene therapy; autoimmune disease;


```
Best Local Similarity 45.6%; Pred. No. 0.11;
Matches 199; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
OY 54 GCGCCACCATACACATCCGCCGCGGCGGCGGCTGCCGCGAGATGACGACCG 113
    |||||
Db 129906 GCGCGTGGCGGGGACCTGCGCGGCGGCGGACGAGCGCGCGCTGCGC 129847
OY 114 TGAATGAGCTTGGGGGGGGGCGACAGACGGCTTGTGATTCGCCCGCGCGCATGGG 173
    |||||
Db 129846 CGCCGTGAGCTGTGTGCGCGGGGCGCTGGCGCGGAGCGCGCTGCT 129787
OY 174 AGCCCTTCAAGCTCAACCCGCGGCGACAGAGCTGTCCGCGGCGAGAGTCCGGCTTAC 233
    |||||
Db 129786 GAGCTCCGCCCGCGCGCGCGCGCGGACCTGTCTTCCAGAACGAGCCTGGCCCTT 129727
OY 234 GTCCGAGGGCCCGCGGCGCTTACCCGCGCTCCGCTGCGGCTGCGCGCGCGAGCGCT 293
    |||||
Db 129726 GGTGGCGGACACCGTGTGCGCGCGCGCGCTGCTGCGCGCGCGCTCGCGCGGGA 129667
OY 294 CGGCGCCCGACGCGCGGCGACGTTGGCTTCTACTCTGTGGGCGCGCGCTTCAACTCCACCGGGA 353
    |||||
Db 129666 GCGCGCGCAAGGCGCAAGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 129607
OY 354 CTTCCTGTCCGCGACGCGCGGCGCTTCCGCGGAGCTTCCGCGCGCGCGCGCGCGCGCG 413
    |||||
Db 129606 CCGCAGAGAGAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129547
OY 414 GCTGTTCGGCGCGGCGCGCGCGCGCTTGCACACGCGGCGACTCGAGCGCGAGGCGCACT 473
    |||||
Db 129546 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129487
OY 474 CCTGTCCCGGCGCG 487
    |||||
Db 129486 GCGCGCGCGAGGCG 129473

RESULT 15
AA199683/c
ID AA199683 standard; DNA: 4403765 BP.
AC AA199683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SRQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI: 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX determining the nucleotide sequence of the strain at positions in the
XX genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ -
XX
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
```

```
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPRO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.
CC
xx
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 11.2%; Score 56; DB 22; Length 4403765;
Best Local Similarity 49.2%; Pred. No. 0.086;
Matches 177; Conservative 0; Mismatches 180; Indels 3; Gaps 1;

OY 43 GCGAGCTTGGCGCGCGCACATACACACTCCGCGCGCGCGCGCGCGCGCGCGCGAG 102
    |||||
Db 3940617 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940558
OY 103 ATCCAGAGACCGTGAACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
    |||||
Db 3940557 CTGCTCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940498
OY 163 GCGCAATGGGAGCGCTTCAAGCTCAACCCGCGCGCGCGCGCGCGCGCGCGCGCGAG 222
    |||||
Db 3940497 GCGGACACCGGGCTTACCGCGCGCGCTCCGCTGCGCTGAAGAGCTGGCGCGCGACAC 3940438
OY 223 TCGGCGTTCAAGTTCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
    |||||
Db 3940437 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940378
OY 283 GCGGACAGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 342
    |||||
Db 3940377 GGTGGGGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940321
OY 343 TCCACCGGGGACTTCCGTGTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 402
    |||||
Db 3940320 GCGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940261
```

Search completed: October 10, 2002, 14:50:58
Job time : 1424 secs

Query Match	28.3%;	Score 141.4;	DB 3;	Length 3923;
Best Local Similarity	66.7%;	Pred. No. 7.9e-18;		
Matches 202; Conservative	0;	Mismatches 101;	Indels 0;	Gaps 0

[illegible]

RESULT 2
 US-09-281-476-20
 Sequence 20, Application US/09281476
 Patent No. 6316597
 GENERAL INFORMATION:
 APPLICANT: Goodman, Peter
 APPLICANT: Koodman, Peter
 TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
 TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Scully, Scott, Murphy & Presser
 STREET: 400 Garden City Plaza
 CITY: Garden City
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 11530
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/281,476
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/860,635
 FILING DATE:
 APPLICATION NUMBER: AU PM9835
 FILING DATE: 05-DEC-1994
 APPLICATION NUMBER: PCT/AU95/00799
 FILING DATE: 29-NOV-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Digilio, Frank S.
 REGISTRATION NUMBER: 31,346
 REFERENCE/DOCKET NUMBER: 10981
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 516-742-4343
 TELEFAX: 516-742-4366
 TELEX:
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3923 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-09-281-476-20

```

Query Match	28.3%	Score 141.4;	DB 4;	Length 3923;
Best Local Similarity	66.7%;	Pred. No. 7.9e-18;		
Matches 202; Conservative	0;	Mismatches 101;	Indels 0;	Gaps 0;

[illegible]

RESULT 3
 US-08-860-635A-18
 : Sequence 18, Application US/08860635A
 Patent No. 6143878
 GENERAL INFORMATION:
 APPLICANT: Koopman, Peter
 APPLICANT: Goodfellow, Peter
 TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
 TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Scully, Scott, Murphy & Presser
 STREET: 400 Garden City Plaza
 CITY: Garden City
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 11530
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,635A
 FILING DATE: 29-MAY-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PM9714
 FILING DATE: 29-NOV-1994
 APPLICATION NUMBER: AU PM9835
 FILING DATE: 05-DEC-1994
 APPLICATION NUMBER: PCT/AU95/00799
 FILING DATE: 29-NOV-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: DiGioglio, Frank S.
 REGISTRATION NUMBER: 31,346
 REFERENCE/DOCKET NUMBER: 10981
 TELECOMMUNICATION INFORMATION:


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MOI 251.5-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS
; LENGTH: 3243 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-468-579B-6
;
Query Match      23.6%; Score 118; DB 2; Length 3243;
Best Local Similarity 38.1%; Fred.No. 1.2e-15;
Matches 208; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
;
QY      82 ACCGACCTGCATCGCCCGGGGCCGCCAGGCCCCCACAAGAACCCTGTGGGCCCCGGCGGG 141
DB      890 ATCAGCCTGACTCATCCCAAGCAGAGGCGCTGGAGAGCGCTGTGTGCACCCACTG 949
QY      142 GCGGGCGCGGGCGGGGGGAGGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGCAAGGCCA 201
DB      950 GAGGAAGCCATGCTGAGCTGCGACATGGATGSGCTCCCGCACCTTCCCAGATCCCAAA 1009
QY      202 CAGGACCGGGTCAAACGGGCCCATGAACGCTTTCAATGATGTGTGTCTCCCGGGGACGGCGCC 261
DB      1010 AGCAGCCACATTCATAAGGGGCCATGAACGCTTCATGATGTGTGGSCCAAAGATGAGAGG 1069
QY      262 AAGATGGCCAGGAGAACCCCAAGATGCACAACCTGGAGATCAGACAAGCCCTGGGGGGCC 321
DB      1070 AAGATCCTTGAAACCTTCCCAAGATGCACAACCTCAGCATCAGACAGATCTTGTGATCT 1129
QY      322 GAGTGAAGAGGTCACTGTCCGAGGCGCAGAAAGGGCGGCTTCATCGACGCGAGGGCCAGCGGCTG 381
DB      1130 CGCTGAAGTCCATGACCAACAGAGAGAACGCGCTCTACTGTAGGAACAAGCGCGGCTG 1189
QY      382 CGCGCGCTGCATGAAGAGCACC CGGATTACAAAGTACCGGGCGCGCCGCAAGACCA 439
DB      1190 AGCGCGCAGCAGCTGGAGAAATCTCTGACTACAAATACAGTACAGCGCGGCCCAAGCGCA 1247
;
RESULT 9
US-08-468-577B-6
; Sequence 6, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Radin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:

```

```

? MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
? COMPUTER: Apple Macintosh
? OPERATING SYSTEM: System 7.5
? SOFTWARE: Wordperfect
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/468,577B
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/239,276
? FILING DATE: 05-MAY-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/872,646
? FILING DATE: 08-JUN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/715,181
? FILING DATE: 14-JUN-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/441,703
? FILING DATE: 04-DEC-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/312,543
? FILING DATE: 17-FEB-1989
? ATTORNEY/AGENT INFORMATION:
? NAME: Kurt G. Briscoe
? REGISTRATION NUMBER: 33,141
? REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (914) 332-1700
? TELEFAX: (914) 332-1844
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3243 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-468-577B-6
?
Query Match          23.6%; Score 118; DB 3; Length 3243;
Best Local Similarity 58.1%; Pred. No. 1.2e-13;
Matches 208; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
?
QY 82 ACCGACCTGACACTCGCGCGGCGGCGGCCCGCCAGAACCTCTGGGCGCGCGCGG 141
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 890 ATCAGCTGTGACTCATCTCCAGCCAGAGCGCGCTGAGAGCGCTGTGTGACCCACTG 949
?
QY 142 GCGGGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 201
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 950 GAGGAGCGCATCTGAGCGCGGCGGAGCATGATGCTCCCGCACTTCCCGAGTCCGGAAC 1009
?
QY 202 CAGGACCGGCGTCAAGAGCGCCATGAAGCGCTTATGATGTGTGTCGCCGCGGCGGCGGCGG 261
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1010 AGAGGACCATATCAAGAGCGCCATGAAGCGCTTATGATGTGTGTCGCCGAGATGAGCGGAGG 1069
?
QY 262 AAGATGGCCAGAGAGAACCCCAAGATGCACAACATCGAGATCAGACAGCGCTGGGGCC 321
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1070 AAGATCTCTCAAGCGCTTCCCAAGATGCACAACATCTCAGATCAGAGATCTTGGATCT 1129
?
QY 322 GAGTGAAGTATCTCCAGGCGGAGAGCGGCGCTTATCATGACGAGGCCAAGCGGCTG 381
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1130 CGCTGGAAGTCCATGACCAACAGAGAGAGCGGCTTACTATGAGAGAACAGCGGCGGCTG 1169
?
QY 382 CGCGGCGTCAATGAGAGAGAGCGGATTAAGTACCGCGCGCGGCGGAGAACCA 439
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1190 AGCGGCGAGACCTGAGAGAGATCTCTGACTACAGTACAGTACAAAGCGGCGGCGGCAAGGCA 1247
?

```

RESULT 10
US-08-239-276-5
; Sequence 5, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel

```

? TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
? TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Sprung Kramer Schaefer & Briscoe
? STREET: 660 White Plains Road
? CITY: Tarrytown
? STATE: New York
? COUNTRY: USA
? ZIP: 10591-5144
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
? COMPUTER: Apple Macintosh
? OPERATING SYSTEM: System 7.5
? SOFTWARE: Wordperfect
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/239,276
? FILING DATE: 05-MAY-1994
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/872,646
? FILING DATE: 08-JUN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/715,181
? FILING DATE: 14-JUN-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/441,703
? FILING DATE: 04-DEC-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/312,543
? FILING DATE: 17-FEB-1989
? ATTORNEY/AGENT INFORMATION:
? NAME: Kurt G. Briscoe
? REGISTRATION NUMBER: 33,141
? REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (914) 332-1700
? TELEFAX: (914) 332-1844
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1570 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-239-276-5
?
Query Match          19.4%; Score 97; DB 2; Length 1570;
Best Local Similarity 56.7%; Pred. No. 7.1e-10;
Matches 198; Conservative 0; Mismatches 150; Indels 1; Gaps 1;
?
QY 82 ACCGACCTGACACTCGCGCGGCGGCGGCCCGCCAGAACCTCTGGGCGCGCGCGG 141
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 38 ATCAGCTGTGACTCATCTCCAGCCAGAGCGCGCTGAGAGCGCTGTGTGACCCACTG 97
?
QY 142 GCGGGCGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 201
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 98 GAGGAGCGCATCTGAGCGCGGCGGAGCATGATGCTCCCGCACTTCCCGAGTCCGGAAC 157
?
QY 202 CAGGACCGGCGTCAAGAGCGCCATGAAGCGCTTATGATGTGTGTCGCCGCGGCGGCGGCGG 261
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 158 AGAGGACCATATCAAGAGCGCCATGAAGCGCTTATGATGTGTGTCGCCGAGATGAGCGGAGG 217
?
QY 262 AAGATGGCCAGAGAGAACCCCAAGATGCACAACATCGAGATCAGACAGCGCTGGGGCC 321
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 218 AAGATCTCTCAAGCGCTTCCCAAGATGCACAACATCTCAGATCAGACAGATCTTGGATCT 277
?
QY 322 GAGTGAAGTATCTCCAGGCGGAGAGCGGCGCTTATCATGACGAGGCCAAGCGGCTG 381
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 278 CGCTGGAAGTCCATGACCAACAGAGAGAGCGGCTTACTATGAGAGAACAGCGCTGTGCTG 337
?
QY 382 CGCGGCGTCAATGAGAGAGAGCGGATTAAGTACCGCGCGCGCGCGC 430
? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 338 ATCG-TCAATCACCCTGAGAGAGATCTCTGACTACAGTACAGTACAAAGCGCGGCGC 385
?

```


LENGTH: 1397 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-579B-1

Query Match 17.0%; Score 84.8; DB 2; Length 1397;
Best Local Similarity 65.3%; Pred. No. 1.1e-07; Mismatches 73; Indels 1; Gaps 1;
Matches 139; Conservative 0; Indels 1; Gaps 1;

QY 218 GGGCCATGAACGCTTCATGTTGTGTCCTCCGGGCGCAGCGCCGCAAGATGCCCCAGAGA 277
DB 1 GGGCCATGAACGCTTCATGTTGTGTCCTCCGGGCGCAGCGCCGCAAGATGCCCCAGAGA 60
QY 278 ACCCCAGATGCACAACTCGAGATCAGCAGCGCTGGGGCCGAGTGGAGGTGATGT 337
DB 61 TCCGAGACATGCACAACTCCAGCATCAGCAGAGATCCTTGATCTCGCTGGAATCCATGA 120
QY 338 CCGAGCGCGAGAGCGGCGCTTCATCGACAGGCCAAGCGGCTGGCGCGCTGCACATGA 397
DB 121 CCAACGAGAGAGAGAGCGCTTCATCGAGAGAGAGCGCTGCTGATCG-TCAATCAGCTGG 179
QY 398 AGAGACACCGGATTCAGATACCGCGCGCGCC 430
DB 180 AGAGTATCTGACTACAGATACAGCGCGCGC 212

RESULT 15
US-08-468-577B-1
Sequence 1, Application US/08468577B
Patent No. 6001804

GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,577B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1397 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-577B-1

Query Match 17.0%; Score 84.8; DB 3; Length 1397;
Best Local Similarity 65.3%; Pred. No. 1.1e-07; Mismatches 73; Indels 1; Gaps 1;
Matches 139; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 218 GGGCCATGAACGCTTCATGTTGTGTCCTCCGGGCGCAGCGCCGCAAGATGCCCCAGAGA 277
DB 1 GGGCCATGAACGCTTCATGTTGTGTCCTCCGGGCGCAGCGCCGCAAGATGCCCCAGAGA 60
QY 278 ACCCCAGATGCACAACTCGAGATCAGCAGCGCTGGGGCCGAGTGGAGGTGATGT 337
DB 61 TCCGAGACATGCACAACTCCAGCATCAGCAGAGATCCTTGATCTCGCTGGAATCCATGA 120
QY 338 CCGAGCGCGAGAGCGGCGCTTCATCGACAGGCCAAGCGGCTGGCGCGCTGCACATGA 397
DB 121 CCAACGAGAGAGAGCGCTTCATCGAGAGAGAGCGCTGCTGATCG-TCAATCAGCTGG 179
QY 398 AGAGACACCGGATTCAGATACCGCGCGCGCC 430
DB 180 AGAGTATCTGACTACAGATACAGCGCGCGC 212

Search completed: October 10, 2002, 17:33:14
Job time : 31.7333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 695.733 Seconds
(without alignments)
9699.805 Million cell updates/sec

Title: US-09-489-101a-4_COPY_1_500
Perfect score: 500
Sequence: 1 ccgcgcgtatgctccagg.....ctcctgagcgcgcgcgcg 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	280.8	56.2	467	9	BE119829 UI-R-CAO-
2	265.6	53.1	739	10	BF305922
3	265.4	52.9	965	9	AI356682
4	264.4	52.9	443	9	AI373018
5	264.4	52.9	450	9	AI052267
6	264.4	52.9	452	9	AI631443
7	264.4	52.9	460	9	AI292258
8	264.4	52.9	460	9	AM005368
9	264.4	52.9	463	9	AI199308
10	264.4	52.9	465	9	AI096977
11	264.4	52.9	465	9	AI146367
12	264.4	52.9	478	10	BF513177
13	264.4	52.9	480	9	AI146406
14	264.4	52.9	483	9	AI421743
15	264.4	52.9	485	9	AM028031
16	264.4	52.9	490	9	AI418146
17	264.4	52.9	507	9	AI571299

18	264.4	52.9	509	9	AI094794
19	264.4	52.9	514	9	AI566261
20	264.4	52.9	537	9	AI971611
21	264.4	52.9	537	9	AM071909
22	264.4	52.9	552	9	AM131791
23	264.4	52.9	580	10	BF939816
24	264.4	52.9	641	9	BE206866
25	264.4	52.9	694	10	BI917204
26	264.4	52.9	728	10	BE672418
27	264.4	52.9	740	10	BE792833
28	264.4	52.9	764	10	BG827056
29	264.4	52.9	834	10	BE797259
30	264.4	52.9	861	10	BE795592
31	264.4	52.9	898	10	BF305566
32	262.8	52.6	427	9	AM615144
33	262.8	52.6	462	9	AA956131
34	262.8	52.6	462	9	AA943207
35	262.8	52.6	462	10	BF512088
36	262.8	52.6	490	9	AA875101
37	262.8	52.6	517	9	AI112078
38	262.8	52.6	537	9	AI480221
39	262.8	52.6	567	9	AI145897
40	262.8	52.6	600	9	AI102567
41	261.2	52.2	460	9	AA924896
42	261.2	52.2	460	9	AA924900
43	261.2	52.2	468	9	BE104368
44	261.2	52.2	472	9	AI097136
45	261.2	52.2	506	9	AM533022

ALIGNMENTS

RESULT 1
LOCUS BE119829 467 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-CAO-ban-e-05-0-UI-s1 UI-R-CAO Rattus norvegicus cDNA clone
UI-R-CAO-ban-e-05-0-UI 3', mRNA sequence.
BE119829
ACCESSION BE119829.1 GI:8511934
VERSION
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 467)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

JOURNAL
MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 76-194, >(CGG
n)Simple_repeat
Seq primer: M13 forward
POLYA=No.

FEATURES
source
Location/Qualifiers

1..467
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAO-ban-e-05-0-UI"

FEATURES	source	1. .739	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4123153" /clone_lib="NIH_MGC_17"
FEATURES	source	1. .739	Location/Qualifiers
FEATURES	source	1. .739	High quality sequence stop: 677.
FEATURES	source	1. .739	found through the I.M.A.G.E. Consortium/LINL at: Image.LINL.gov
FEATURES	source	1. .739	clone distribution: MGC clone distribution information can be
FEATURES	source	1. .739	DNA sequencing by: Incyte Genomics, Inc.
FEATURES	source	1. .739	CDNA library arrayed by: The I.M.A.G.E. Consortium (LINL)
FEATURES	source	1. .739	CDNA library preparation: Ling Hong/Rubin Laboratory
FEATURES	source	1. .739	TISSUE Procurement: ATCC
FEATURES	source	1. .739	Email: ggapds@mail.nih.gov
FEATURES	source	1. .739	Contact: Robert Strausberg, Ph.D.
FEATURES	source	1. .739	Unpublished (1999)
FEATURES	source	1. .739	National Institutes of Health, Mammalian Gene Collection (MGC)
FEATURES	source	1. .739	NIH-MGC http://mgc.nci.nih.gov/
FEATURES	source	1. .739	1 (bases 1 to 739)
FEATURES	source	1. .739	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
FEATURES	source	1. .739	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
FEATURES	source	1. .739	human.
FEATURES	source	1. .739	BF305922.1 GI:11252955
FEATURES	source	1. .739	EST.
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES	source	1. .739	mRNA sequence.
FEATURES	source	1. .739	6018892061 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4123153 5',
FEATURES	source	1. .739	BF305922
FEATURES</			

[illegible]

[illegible]

Db	366	CGCCTCATGACAGAGATTAAGTACACGCTGCGCGGCGGCTGCTGCTGCGCGCGC	419
RESULT#	8		
LOCUS	AM005368		
DEFINITION	Wz55h12.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2565671.3	460 bp	linear
ACCESSION	AM005368		
VERSION	AM005368.1	GI:5854146	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 460)		
TITLE	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project		
JOURNAL	(CGAP/BRGAP), Tumor Gene Index		
COMMENT	Unpublished (1998) Contact: Robert Strausberg, Ph.D. Email: cgapbs-rt@mail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www-bio.ln1l.gov/dbfp/image/image.html Insert length: 1157 Std Error: 0.00 Seq primer: -40mp from Gibco High quality sequence stop: 457. Location/Qualifiers 1. 460 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2565671" /clone_lib="NCI CGAP_Brn25" /tissue_type="anaplastic oligodendroglioma" /lab_host="DH10B" /note="Organ: brain; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTCCATCTGAGAGTGCGAGCGCGCCGACATAGGTTTTTTTTTTTTTTT TTTTTCATCATGAGTGGATGCTTTCATGTTGTGTCCGCGGCGCGCAACCGCA 185 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."		
FEATURES	source		
BASE COUNT	104 a 140 c 165 g 51 t		
ORIGIN			
Query Match	52.9%; Score 264.4; DB 9; Length 460;		
Best Local Similarity	84.2%; Pred. No. 2.3e-26;		
Matches 298; Conservative	0; Mismatches 56; Indels 0; Gaps 0;		
0Y	143 CGGGGGGGGGGGGGGAGGCGGGGGGGGGGGGGGGGGGGGGCGGCAAGGCCAACCC 202		
Db	66 CGGGGGGGGGGGGGGCACTCACCGCGGGCGGGCGGGCGGCAACGAAAAACAGCC 125		
0Y	203 AGACCGGGGTCAACGGCCCATGAAACGCTTCATGGTGTGTCCCGGGCGAGCGGCGCA 262		
Db	126 CGGACCGCGCTCAAGGGGGCCCATGAAATGCTTCATGGTGTGTGTCCCGGGCGAGCGGCGCA 185		
0Y	263 AGATGGCCCGAGAGAACCCCAAGATGCAACAACCTGGAGATGACGAAGCGCTGGGGGCGCC 322		
Db	186 AGATGGCCCGAGAGAACCCCAAGATGCAACAACCTGGAGATGACGAAGCGCTGGGGGCGCC 245		
0Y	323 AGTGAAGTCAATGTCGAGGCCGAGAGAGCGGCGCTTCATGCAGAGGCGCAAGCGGCTGC 382		

[illegible]

Db	271	AGTGGAAACCTTTTGGCGGAGACGGAGGAGCGCGCCCTTCATCGACGAGGCTAAAGCGCGTGC	330
0y	383	CGCGGCTCGACATGAAAGAGACACCGGATTTACAACTACCGCGCGCGCGCAAGACCAAGA	442
Db	331	GAGCGCTCGACATGAAAGAGACACCGGATTTAAATATACCGGCGCGCGGAGAAACCAAGA	390
0y	443	CGCTGCTCAAGAGACAACTACTGCTGCGCGCGCGCGGCTCTGCGCGCGCGGCGG	496
Db	391	CGCTCATGAAAGAGATTAAGTACACGCTGCGCGCGCGGCTCTGCGCGCGCGGCGG	444
RESULT 15			
AM028031			
LOCUS			
DEFINITION	AM028031	485 bp	linear
ACCESSION	AM028031		EST 08-MAR-2000
VERSION	AM028031.1		
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
SOURCE			
BASE COUNT			
ORIGIN			
Query Match	52.9%	Score 264.4	DB 9; Length 485;
Best Local Similarity	84.2%	Prod. No. 2.3e-26;	
Matches 298; Conservative	0;	Mismatches 56;	Indels 0; Gaps 0;
0y	143	CGGCGCGCGCGCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCAAGCGCAAC	202
Db	45	CGGCGCGCGCGCGCGCGCAATTCACCGCGCGCGCGCGCGCGCGCGCAACGAAAAACACCC	104
0y	203	AGGACGGGATCAACGCGCCATGAAGCGCTTCATGCTGTGCTCCCGCGGCGACGCGGCGCA	262

```

Db      105  CGGACCGCGTCAGCGGCGCCATGATGCTTCATGTTGTGTCCCGCGGGCAGCGGCGCA 164
QY      263  AGATGGCCCCAGAGAACCCCAAGATGCACAACTCGAGATCAGCAGCGCCTGGGGCGCG 322
Db      165  AGATGGCCCCAGAGAACCCCAAGATGCACAACTCGAGATCAGCAGCGCCTGGGGCGCG 224
QY      323  AGTGAAGATCATGTCCGAGGCGGAGAGGCGGCTTCATCGAGAGGCCCAAGCGGCTGC 382
Db      225  AGTGAAGATCATGTCCGAGGCGGAGAGGCGGCTTCATCGAGAGGCCCAAGCGGCTGC 284
QY      383  GCGCGCTGCACATGAGAGGAGCACCAGATTACAACTACCGGCGCGCGCAGACCAAGA 442
Db      285  GAGCGCTGCACATGAGAGGAGCACCAGATTATAATACCGGCGCGCGGAGAAACCAAGA 344
QY      443  CGCTGCTCAAGAGAGCACTACTCGCGCGGCGGCTCTGTGCGCGCGCGCG 496
Db      345  CGCTCATGAAGAGATAGTACACGCTGCGCGGCGGCTGTGCGCGCGCGCG 398

```

Search completed: October 10, 2002, 20:27:16
 Job time : 698.733 secs

XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
 XX WPI; 2001-457597/49.
 DR P-PSDB; AAE05811.
 XX
 PT Isolated polypeptide, used to treat or prognose a disorder
 PT characterized by expression of a hCAP e.g. cancer, is encoded by an
 PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
 XX
 PS Claim 57; Page 88-90; 152pp; English.
 XX
 CC The invention relates to nucleic acids and encoded polypeptides which
 CC are cancer associated antigens expressed in patients afflicted with
 CC small cell lung cancer. The molecules provided by the invention can be
 CC used in the diagnosis, monitoring, research or treatment of conditions
 CC characterized by the expression of one or more cancer associated
 CC antigens. The polypeptide is used to treat a disorder characterized by
 CC expression of a hCAP, and determine regression, progression or onset
 CC of a condition characterized by expression of an abnormal amount of a
 CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
 CC are small and non-small cell lung cancer, melanoma, colon, breast, head
 CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
 CC The present sequence is a small cell lung cancer associated gene
 CC designated as NY-SCLC-2, encoding human SOX1 protein.
 XX
 SQ Sequence 4091 BP; 924 A; 1109 C; 1110 G; 943 T; 5 other;
 Query Match 100.0%; Score 500; DB 22; Length 4091;
 Best Local Similarity 100.0%; Pred. No. 9.6e-70;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCGCTATGCTCCAGGCGCTCTCTCCGCGGCGGTGATACCCGCGACCGCCCG 60
 DB 1 CCGGCGCTATGCTCCAGGCGCTCTCTCCGCGGCGGTGATACCCGCGACCGCCCG 60
 QY 61 ATGTACAGCATGATGATGAGACCGACCTGCTCCGCGCGCGCCGACGCGCCGACG 120
 DB 61 ATGTACAGCATGATGATGAGACCGACCTGCTCCGCGCGCGCCGACGCGCCGACG 120
 QY 121 AACCTCGGCG 180
 DB 121 AACCTCGGCG 180
 QY 181 GCGGCGGCG 240
 DB 181 GCGGCGGCG 240
 QY 241 TGGTCCCG 300
 DB 241 TGGTCCCG 300
 QY 301 ATAGAGCAAGGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 DB 301 ATAGAGCAAGGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 QY 361 ATGAGAGAGGCG 420
 DB 361 ATGAGAGAGGCG 420
 QY 421 CGGCG 480
 DB 421 CGGCG 480
 QY 481 CTCTGCG 500
 DB 481 CTCTGCG 500

RESULT 2
 AAX16151
 ID AAX16151 standard; cDNA to mRNA; 1542 BP.
 XX

AC AAX16151;
 XX
 DT 22-APR-1999 (first entry)
 XX
 DE Human Sox1 encoding cDNA.
 XX
 KW Human; Sox1; neuronal stem cell gene; neuroblastic cell; cancer;
 KW nervous system; neurological disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT misc_difference 1..60 /tag= a
 FT /note= "specifically claimed Sox1 control sequence"
 FT CDS 60..1223 /tag= b
 FT /product= "Sox1"
 XX
 XX W09900516-A2.
 XX
 PD 07-JAN-1999.
 XX
 PF 25-JUN-1998; 98WO-GH01862.
 XX
 XX 25-JUN-1997; 97GB-0013469.
 XX
 XX (MED1-) MEDICAL RES COUNCIL.
 XX
 PA Lovell-Badge R, Pevery LH, Smith A;
 PI WPI; 1999-095759/08.
 XX
 DR P-PSDB; AAM94461.
 XX
 XX
 XX Isolating neuroblastic cells from a population - by detecting the
 PT expression of the Sox1 gene in the cells and sorting the cells to
 PT isolate those cells expressing Sox1
 XX
 PS Claim 13; Page 45-47; 60pp; English.
 XX
 CC A method has been developed for isolating neuroblastic cells from a cell
 CC population. The method comprises: (a) detecting Sox1 gene expression in
 CC the cells; and (b) isolating those cells expressing Sox1. Also described
 CC is a method for producing a cell committed to the neuronal lineage,
 CC comprising: (a) transfecting a pluripotent stem cell with a genetic
 CC construct encoding Sox1 expression; (b) culturing the stem cells to
 CC differentiate into neural cells; and (c) isolating those neural cells
 CC produced. The present sequence encodes human Sox1. Detection of Sox1
 CC expressing cells is important in diagnosing and treating cancers of the
 CC nervous system. Neural stem cells are useful for the treatment of
 CC neurological disorders, especially for repair of accidentally induced
 CC trauma in the CNS or for correction of congenital or pathological
 CC diseases of the CNS. A patient with a neurological disorder can act as a
 CC self-donor. Cells may be isolated from the patient and either sorted to
 CC extract neuroblasts, or treated in order to differentiate neuroblasts,
 CC from specific or general precursors.
 XX
 SQ Sequence 1542 BP; 238 A; 565 C; 560 G; 179 T; 0 other;
 Query Match 97.6%; Score 488; DB 20; Length 1542;
 Best Local Similarity 99.8%; Pred. No. 7.8e-68;
 Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CCGGCGCTATGCTCCAGGCGCTCTCTCCGCGGCGGTGATACCCGCGACCGCCCGG 60
 DB 1 CCGGCGCTATGCTCCAGGCGCTCTCTCCGCGGCGGTGATACCCGCGACCGCCCGG 59
 QY 61 ATGTACAGCATGATGATGAGACCGACCTGCTCCGCGCGCGCGCGCGCGCGCGCG 120
 DB 61 ATGTACAGCATGATGATGAGACCGACCTGCTCCGCGCGCGCGCGCGCGCGCGCG 119
 QY 121 AACCTCGGCG 180
 DB 121 AACCTCGGCG 180

FT	variation	/note="replace(322..324, "cgn")
FT		328..330
FT		/*tag= u
FT	variation	/note="replace(328..330, "cun")"
FT		361..363
FT		/*tag= v
FT	variation	/note="replace(361..363, "cgn")"
FT		367..369
FT		/*tag= w
FT	variation	/note="replace(367..369, "cgn")"
FT		370..372
FT		/*tag= x
FT	variation	/note="replace(370..372, "cgn")"
FT		385..387
FT		/*tag= y
FT	variation	/note="replace(385..387, "cun")"
FT		388..390
FT		/*tag= z
FT	variation	/note="replace(388..390, "cun")"
FT		406..408
FT		/*tag= aa
FT	variation	/note="replace(406..408, "agy")"
FT		409..411
FT		/*tag= ab
FT	variation	/note="replace(409..411, "cun")"
FT		421..423
FT		/*tag= ac
FT	variation	/note="replace(421..423, "cun")"
FT		424..426
FT		/*tag= ad
FT	variation	/note="replace(424..426, "cun")"
FT		505..507
FT		/*tag= ae
FT	variation	/note="replace(505..507, "cgn")"
FT		508..510
FT		/*tag= af
FT	variation	/note="replace(508..510, "cun")"
FT		513..515
FT		/*tag= ag
FT	variation	/note="replace(513..515, "agy")"
FT		583..585
FT		/*tag= ah
FT	variation	/note="replace(583..585, "agy")"
FT		631..632
FT		/*tag= ai
FT	variation	/note="replace(631..632, "cun")"
FT		676..678
FT		/*tag= aj
FT	variation	/note="replace(676..678, "cgn")"
FT		742..744
FT		/*tag= ak
FT	variation	/note="replace(742..744, "cgn")"
FT		760..762
FT		/*tag= al
FT	variation	/note="replace(760..762, "cun")"
FT		769..771
FT		/*tag= am
FT	variation	/note="replace(769..771, "agy")"
FT		778..780
FT		/*tag= an
FT	variation	/note="replace(778..780, "agy")"
FT		784..786
FT		/*tag= ao
FT	variation	/note="replace(784..786, "agy")"
FT		799..801
FT		/*tag= ap
FT	variation	/note="replace(799..801, "agy")"
FT		805..807
FT		/*tag= aq
FT	variation	/note="replace(805..807, "agy")"
FT		811..813
FT		/*tag= ar
FT	variation	/note="replace(811..813, "agy")"

[illegible]

```
DB 181 UCGUCNAGRGCNCARAGRAAARAUGCNCARGAARAACCNARAUGCAVAYUUNCAR 240
QY 301 ATCAGCAAGGCGCTGGGGCGAGTGAAGTCTATGTCGCCAGCGCGAGAACGCCCTTC 360
DB 241 AHHUCNAARGRURUGGNGARUCGAAKRUNAUGUCNGARGCNGAANAARAGCCNUY 300
QY 361 ATCGAGAGGCCAAGCGGCTGCCGCTGCACATGAAGAGCACCGGATTAAGTAC 420
DB 301 AHHGAYGARGCNAARAGRURBRAGRCNUURCAVAYUGAARARCAVCCNGAYUAYAAARAY 360
QY 421 CGGGCCCGCCGCAAGCAAGACGCTCTCAAGAAGACAACTACTGCTGGCCGGCGGG 480
DB 361 AGRCACNAGRRARACNARACNUURURARAARAYARUAYUUCNUURGCGNGNGN 420
QY 481 CTCCTGGCGCGCGCGCGG 500
DB 421 UURURGCGNGCGNGCG 440

RESULT 6
AADI1110
ID AADI1110 standard; DNA; 1085 BP.
XX
XX AADI1110;
XX
XX 24-SEP-2001 (first entry)
XX
XX Human small cell lung cancer associated gene, SOX2.
XX
XX Human; small cell lung cancer; therapy; hCAP; nucleic acid; NA;
XX melanoma; cancer; colon; breast; head; neck; transitional cancer;
XX leiomyosarcoma; synovial sarcoma; cytostatic; Sox2; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 13..966
XX FT /*tag= a
XX FT /product= "Human SOX2 protein"
XX
XX W020015349-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US02015.
XX
XX 21-JAN-2000; 2000US-0489101.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX PA (SLOK ) SLOAN KETTERING INST CANCER RES.
XX PA (CORR ) CORNELL RES FOUND INC.
XX
XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y.
XX
XX WPI: 2001-457597/49.
XX P-PSDB; AAEO5810.
XX
XX Isolated polypeptide, used to treat or prognose a disorder
XX characterized by expression of a hCAP e.g. cancer, is encoded by an
XX isolated nucleic acid comprising an NA Group 3 or 4 molecule -
XX
XX Claim 57; Page 87; 152pp; English.
XX
XX The invention relates to nucleic acids and encoded polypeptides which
XX are cancer associated antigens expressed in patients afflicted with
XX small cell lung cancer. The molecules provided by the invention can be
XX used in the diagnosis, monitoring, research or treatment of conditions
XX characterised by the expression of one or more cancer associated
XX antigens. The polypeptide is used to treat a disorder characterised by
XX expression of a hCAP, and determine regression, progression or onset
XX of a condition characterised by expression of an abnormal amount of a
XX protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
XX are small and non-small cell lung cancer, melanoma, colon, breast, head
```

```
CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
CC The present sequence is a small cell lung cancer associated gene
CC designated as NY-SCLC-1, encoding human Sox2 protein.
XX
SQ Sequence 1085 BP; 253 A; 348 C; 348 G; 136 T; 0 other;

Query Match 52.9%; Score 264.4; DB 22; Length 1085;
Best Local Similarity 84.2%; Pred. No. 4, 1e-33;
Matches 298; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 143 CGGGCGCGCGCGGGCGAGGCGGGGCGCGCGCGCGCGCGCGCGCGCGCAAGCCACC 202
DB 65 CGGGCGCGCGCGCGCGCGCACTCCACCGCGCGCGCGCGCGCGCGCGCAAGCCACC 124
QY 203 AGGACCGGGTCAACGGCGCCATGAAACGCTTCACTGTTGTTCCCGCGGCGCGCA 262
DB 125 CGGACCGGGTCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 184
QY 263 AGATGGCCGAGAGAACCCCAAGATGCACACTCGAGATCAGCAAGCGCGCGCGCG 322
DB 185 AGATGGCCGAGAGAACCCCAAGATGCACACTCGAGATCAGCAAGCGCGCGCGCG 244
QY 323 AGTGAAGGTCACTGTCCGAGGCGGAGAGCGGCGCGCGCGCGCGCGCGCGCG 382
DB 245 AGTGAAGGTCACTGTTCGAGGCGGAGAGCGGCGCGCGCGCGCGCGCGCGCG 304
QY 383 GCGCGCTGCACATGAAGAGCACCGCGATTACAGTACCGCGCGCGCGCGCGCGCA 442
DB 305 GAGCGCTGCACATGAAGAGCACCGCGATTATTAATACCGCGCGCGCGCGCGCGCA 364
QY 443 CGCTCTCAAGAGAGCAAGTACTGCTGGCGCGCGCGCGCGCGCGCGCGCGCG 496
DB 365 CGCTCTCAAGAGAGTAAAGTACAGCTGCCGCGCGCGCGCGCGCGCGCGCGCG 418

RESULT 7
AAK52388
ID AAK52388 standard; cDNA; 1087 BP.
XX
XX AAK52388;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 933.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX W0200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wehrman T, Goodrich R;
```



```

RESULT 9
AAK52284
ID AAK52284 standard; cDNA; 1126 BP.
XX
AC AAK52284;
XX
DT 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 829.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR P-PSDB; AAM79151.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 2766-2767; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1126 BP; 262 A; 361 C; 364 G; 139 T; 0 other;

Query Match 52.9%; Score 264.4; DB 22; Length 1126;
PS Best Local Similarity 84.2%; Pred. No. 4.1e-33;
Matches 298; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

```

DB 158 CGAGCGGCTCAAGCGGCCCATGATGATCCTTCATGTGTGTGTCCCGGGGACGCGGCA 217
QY AGATGGCCCGAGAGAACCCCAAGATCGACACTCGGAGATCAGACGCGCTGGGGCCG 322
DB 218 AGATGGCCCGAGAGAACCCCAAGATCGACACTCGGAGATCAGACGCGCTGGGGCCG 277
QY AGTGAAGGTCATGTCCGAGCGCCGAGAACGCGCGCTTCATCGACGAGAGCCGAGCTGC 382
DB 278 AGTGAAGGTCATGTTCGCGAGAGCGAGAACGCGCGCTTCATCGACGAGAGCTGC 337
QY 363 GCGGCGCTGCACATGAGAGACGACCCCGATTACAGTACCGCGCGCCCGCAGACAGAGA 442
DB 338 GAGCGCTGCACATGAGAGAGCACCGCGATTATATAATCCGCGCCCGCGGAAAAACAGGA 397
QY 443 CGCTGCTCAGAGAGACGACGACTCGCTGCGCGCGCGGCTCCGCGCGCGGCG 496
DB 398 CGCTCATGAGAGAGATATAGTACAGCGCTCCCGCGCGGCTGCTGGCCCGCGGCG 451

RESULT 10
AAK53372
ID AAK53372 standard; cDNA; 1501 BP.
XX
AC AAK53372;
XX
DT 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 2901.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR P-PSDB; AAM80239.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 5070; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and

```

CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM60020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1501 BP; 391 A; 428 C; 450 G; 220 T; 12 other;
Query Match 52.9%; Score 264.4; DB 22; Length 1501;
Best Local Similarity 84.2%; Pred. No. 4e-33;
Matches 298; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
OY 143 CGGCGCGCGCGGGGCGGAGCGGGGCGGGGCGGGGCGGCGCAAGGCCAAC 202
DB 91 CGGGGGGGGGGGGGGCACTCCACCGCGCGCGCGCGCAACCAAGAAAGGCC 150
OY 203 AGACACGGGCTCAACGGCCCATGAACGCTTCATGTGTGTCCCGCGGCGAGCGCGCA 262
DB 151 CGGACCGCGTCAACGGCGCCCATGAATGCTTCATGTGTGTCCCGCGGCGAGCGCGCA 210
OY 263 AGATGGCGCCAGGAACCCCAAGATGCACAACTCGAGATCAGCAAGCCCTGGGGCGCG 322
DB 211 AGATGGCGCCAGGAACCCCAAGATGCACAACTCGAGATCAGCAAGCCCTGGGGCGCG 270
OY 323 AGTGAAGTGTATGCTCCGAGCGCGGCGGCGGCTTCATGCGAGCGCAAGCGGCTGC 382
DB 271 AGTGAAGTGTATGCTCCGAGCGCGGCGGCGGCTTCATGCGAGCGCAAGCGGCTGC 330
OY 383 GCGCGCTGCACATGAAGAGCACCGCGATTAACAGTACCGCGCGCGCGCAAGACCAGA 442
DB 331 GAGCGCTGCACATGAAGAGCACCGCGATTAACAGTACCGCGCGCGCGCAAGACCAGA 390
OY 443 GCGCTGCACGAAGAAGACAGTACTGCTGGCGCGCGGCTCTGGCGCGCGCGG 496
DB 391 GCGCTGCACGAAGAAGATGATACAGCTGCCGCGGCTGCTGGCGCGCGG 444
RESULT 11
ABI99318
ID ABI99318 standard; cDNA; 2283 BP.
XX
AC ABI99318;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:179.
XX
KW Mouse; ischemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
PN M0200188188-A2.
XX
PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-JP04192.
XX
PR 18-MAY-2000; 2000JP-0145977.
XX
PA (UNINT-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takehashi Y, Nagata T, Ishii Y;
XX
DR WPI; 2002-034733/04.
XX
DR P-PSDB; ABB57082.
XX
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX

PS Claim 2; Page 518-521; 2690pp; English.
XX
XX The present invention describes a method for examining ischemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischemic condition (e.g. compressive
CC ischemia, occlusive ischemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 2283 BP; 597 A; 583 C; 614 G; 488 T; 1 other;
Query Match 52.6%; Score 262.8; DB 24; Length 2283;
Best Local Similarity 82.9%; Pred. No. 6.7e-33;
Matches 300; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
OY 137 CGGGGGCGGGCGCGCGCGCGGAGCGGGGCGGGGCGGGGCGGCGCAAG 196
DB 414 CGGGGGCGGGCGGGGAGAGAGCAACCGCGCGCGCGCGCAACAGAA 473
OY 197 CCAACGAGACCGGGTCAACCGCCCATGAACGCTTCATGTGTGTCCCGGGCAGC 256
DB 474 ACAGCCCGGACCGCGTCAAGAGGCCCATGAAGCGCTTCATGTGTGTCCCGGGCAGC 533
OY 257 GGGCAGATGGGCCCAAGAGAACCCCAAGATGCACAACTCGAGATCAGCAAGCGCTGC 316
DB 534 GCGCTTAAGATGGGCCAAGAGAACCCCAAGATGCACAACTCGAGATCAGCAAGCGCTGC 593
OY 317 GGGCGGAGTGAAGTATGTCGAGGCGGAGAAAGCGGCGCTTCATCAGCAGAGCCAAGC 376
DB 594 GCGCGAGTGAAGACTTTTGTCCGAGACGAGAAAGCGGCGCTTCATCAGCAGAGCCAAGC 653
OY 377 GCGTGGCGGCGCTGCACATGAAGAGCACCGCGATTAACAGTACCGCGCGCGCGCAAGA 436
DB 654 GCGTGGCGGCGCTGCACATGAAGAGCACCGCGATTAACAGTACCGCGCGCGCGCAAGA 713
OY 437 CCAAGAGCTGCTCAAGAGAGCAAGTACTGCTGGCGGGGCGGCTCTGGCGCGCGCGG 496
DB 714 CCAAGAGCTGCTCAAGAGAGATGATACAGCTTCCGGAGGCTTGGTGGCCCCCGGCG 773
OY 497 CG 498
DB 774 GG 775
RESULT 12
AAD1118
ID AAD1118 standard; DNA; 2509 BP.
XX
AC AAD1118;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human small cell lung cancer associated gene, SOX3.
XX
KW Human; small cell lung cancer; therapy; hCAP; nucleic acid; NA;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW leiomyosarcoma; synovial sarcoma; cytosolic; SOX3; ds.
XX
OS Homo sapiens.
XX
XX
FH Key location/Qualifiers
FT CDS 441..1773
FT /*tag= a
FT /product= "Human SOX3 protein"
XX

XX
PA (CORI-) CORIXA CORP.
XX
PI Lodes MJ, Wang T, Mohamach R, Indirias CY;
XX
DR WPI: 2002-010896/01.
XX
PT Lung tumour polynucleotide and polypeptides useful in therapy and
XX diagnosis of cancer especially lung cancer
PS
PS Claim 1; Page 180; 295pp; English.
XX
CC The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by
CC utilising oligonucleotides (III), where the biological sample
CC from the patient is contacted with (III), detecting the amount of
CC polynucleotide hybridised to (III) in the sample and comparing the
CC amount of polynucleotide to a predetermined cut-off value and thereby
CC determining cancer in a patient. (I), (II) or antigen-presenting cells
CC expressing (II) is useful for stimulating and/or expanding T cells
CC specific for a tumour protein. The method comprises contacting T cells
CC with one of the components under conditions to permit the stimulation
CC and/or expansion of the cells. A composition comprising (I) is useful for
CC stimulating an immune response in a patient and for inhibiting the
CC development of a cancer especially lung cancer in a patient. An
CC isolated T cell population is useful for removing tumour cells from the
CC biological sample and for inhibiting the development of cancer in a
CC patient. AAS61460-AAS61874 represent novel human lung small cell
CC cancer antigen coding sequences of the invention.
XX
SQ Sequence 412 BP; 83 A; 133 C; 144 G; 51 T; 1 other;
XX
Query Match 44.2%; Score 220.8; DB 24; Length 412;
Best Local Similarity 84.1%; Pred. No. 2.7e-26;
Matches 249; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 204 GGACGGGTCAACGGGCCCTTCATGATGTGTGTCCCGCGGCAAGCGCGCAA 263
DB 15 GGACGACGTCAAGCGGCCCATGACGCTTCATGTGTGTGTCCCGCGGCTCAGCGCGCAA 74
QY 264 GATGCCCCAGGAGAACCCCAAGATGCAACAATCGAGATCAGCAAGCGCTGGGGGCCGA 323
DB 75 GATGGCCAGGAGAACCCCAAGATGCAACAATCGAGATCAGCAAGCGCTGGGGGCCGA 134
QY 324 GTGGAAGGTCAATGTCGAGCGGAGGAGCGCCGTCATGACAGAGGCCAAGCGGCTCG 383
DB 135 GTGGAAGGTCAATGTCGAGCGGAGGAGCGCCGTCATGACAGAGGCCAAGCGGCTCAG 194
QY 384 CGCGCTGCACATGAAGGAGCACCAGGATTACAAGTACCGGCGCGCCCAAGACCAAGAC 443
DB 195 CGCATGTGACATGAAGGAGCACCAGGATTACAAGTACCGGCGCGCCCAAGACCAAGAC 254
QY 444 GCTGCTCAAGAAAGACAACTACTCGCTGGCGGCGGCGCTCTGGCGGCGGCGCGG 499
DB 255 GCTGCTCAAGAAAGACAACTACTCGCTGGCGGCGGCGCTCTGGCGGCGGCGGCGG 310

Search completed: October 10, 2002, 14:28:39
Job time : 91 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 18.7333 Seconds
(without alignments)
6556.054 Million cell updates/sec

Title: US-09-489-101A-3_COPY_1_500
Perfect score: 500
Sequence: 1 cacacgcgcgcgcgtatcacaa.....atgcacgctacgcgcacat 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	140	28.0	3923	3	US-08-860-635A-20
2	140	28.0	3923	4	US-09-281-476-20
3	134.6	26.9	1661	1	US-08-196-016A-1
4	129.8	26.0	1667	1	US-08-196-016A-2
5	129.4	25.9	2249	3	US-08-860-635A-18
6	129.4	25.9	2249	4	US-09-281-476-18
7	99.8	20.0	3243	2	US-08-239-276-6
8	99.8	20.0	3243	2	US-08-468-579B-6
9	99.8	20.0	3243	2	US-08-468-579B-6
10	89.4	17.9	5051	2	US-08-239-276-2
11	89.4	17.9	5051	2	US-08-468-579B-2
12	89.4	17.9	5051	2	US-08-468-579B-2
13	80.8	16.2	1570	2	US-08-239-276-5
14	80.8	16.2	1570	2	US-08-468-579B-5
15	80.8	16.2	1570	2	US-08-468-579B-5
16	73.6	14.7	1397	2	US-08-239-276-1
17	73.6	14.7	1397	2	US-08-468-579B-1
18	73.6	14.7	1397	3	US-08-468-579B-1
19	62.6	12.5	684	4	US-08-998-416-672
20	49	9.8	7218	1	US-08-332-463-14
21	44.6	8.9	2040	2	US-08-821-355A-1
22	44.6	8.9	2040	2	US-09-003-687A-1
23	44.6	8.9	2040	3	US-09-136-605-1
24	44.6	8.9	2444	2	US-08-821-355A-2
25	44.6	8.9	2444	2	US-09-003-687A-2
26	44.6	8.9	2444	2	US-09-136-605-2
27	44.6	8.9	23673	4	US-09-773-816-1

28	43	8.6	1898	1	US-07-814-964-8	Sequence 8, Appl
29	43	8.6	1898	1	US-08-258-442-8	Sequence 8, Appl
30	43	8.6	1898	1	US-08-328-809-3	Sequence 3, Appl
31	43	8.6	1898	5	PCT-US92-11107-8	Sequence 6, Appl
32	43	8.6	2839	1	US-07-814-964-6	Sequence 6, Appl
33	43	8.6	2839	1	US-08-258-442-6	Sequence 6, Appl
34	43	8.6	2839	1	US-08-328-809-1	Sequence 1, Appl
35	43	8.6	2839	3	US-09-015-003-1	Sequence 9, Appl
36	43	8.6	2839	5	PCT-US92-11107-6	Sequence 6, Appl
37	42.8	8.6	2032	4	US-09-045-284A-1	Sequence 14, Appl
38	42.8	8.6	4722	4	US-08-979-608A-14	Sequence 14, Appl
39	41.6	8.3	1444	1	US-07-814-964-9	Sequence 9, Appl
40	41.6	8.3	1444	1	US-08-258-442-9	Sequence 9, Appl
41	41.6	8.3	1444	1	US-08-328-809-4	Sequence 4, Appl
42	41.6	8.3	1444	5	PCT-US92-11107-9	Sequence 9, Appl
43	41.4	8.3	2277	1	US-08-676-967-5	Sequence 5, Appl
44	41.4	8.3	2277	1	US-08-676-974-5	Sequence 5, Appl
45	41.4	8.3	2277	2	US-09-098-487-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-860-635A-20
Sequence 20, Application US/08860635A
Patent No. 6143878
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,635A
FILING DATE: 29-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3923 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-860-635A-20

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
POSITION IN GENOME:
CHROMOSOME/SEGMENT: Y-chromosome, SRY gene
US-08-196-016A-1

Query Match 26.9%; Score 134.6; DB 1; Length 1661;
Best Local Similarity 69.7%; Pred. No. 2.1e-21;
Matches 182; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 114 GAAACACGCCGCGGACCGGTCACAGCGGCCCTTCATGCTGTCGCCGCGG 173
DB 1052 GGAGGCGCGGACGAGTGGCCGACGAGAACCCCAAGATGCACAACTCGAGATCAAGCG 233
OY 174 GCAGGCGCGGACGAGTGGCCGACGAGAACCCCAAGATGCACAACTCGAGATCAAGCG 233
DB 1112 ACGAAGACGAAAGGTGGCTCTAGAGATCCCAAAATGAAAACTCAGACATCAAGCA 1171
OY 234 CCGTGGCCCGGAGTGAACCTTTGTCGAGACGAGAACGCGCGCTTCATCGACGAGCG 293
DB 1172 GCTGGATATGATGAGTGAAGAGGCTTACAGATGCTGAAAGCGCCCAATCTTGGAGAGCG 1231
OY 294 TAAGCGGCTGCGAGCGCTGCACATGAGAGACACCCGGATTAATAATACCGCGCGCGCG 353
DB 1232 ACAGAGACTACTAGCCATACACCGAGACAAATACCGGCTATTAATATCGACTCGTCG 1291
OY 354 GAAACCAAGACGCTCATGAA 374
DB 1292 GAGAGCAAGAGCGCACAGAA 1312

RESULT 4
US-08-196-016A-2
Sequence 2, Application US/08196016A
Patent No. 5596089

GENERAL INFORMATION:
APPLICANT: SILVERSIDES, David W.
APPLICANT: DANEAU, Isabelle M.F.
APPLICANT: HOUEU, Alain
TITLE OF INVENTION: SRY GENOMIC SEQUENCES AND USE IN TISSUE
TITLE OF INVENTION: SEXING
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: KLAUBER & JACKSON
STREET: Continental Plaza, 411 Hackensack Avenue
CITY: Hackensack
STATE: N.J.
COUNTRY: U.S.A.
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/196, 016A
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JACKSON, David A.
REGISTRATION NUMBER: 26,742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1667 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
POSITION IN GENOME:
CHROMOSOME/SEGMENT: Y-chromosome, SRY gene
US-08-196-016A-2

Query Match 26.0%; Score 129.8; DB 1; Length 1667;
Best Local Similarity 70.6%; Pred. No. 2.4e-20;
Matches 173; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

OY 122 GCCCGACCGCGTCGAAGCGGCCCATGAAATGCTTCATGCTGTCGCCGCGGACGCGG 181
DB 791 GCCAGATCGTGTCAAGGACCATGAAAGCGCTTATGTCGTCTCTGATCAAGA 850
OY 182 GCAAGATGCCCGAGAGAACCCCAAGATGCACACTCGGAGATCAGCAAGCGCTGGGG 241
DB 851 GAAAGTGGCTTAGAGAGACCTCAAAATGCAAACTCAGAGATCAGCAAGTGGCTGGAT 910
OY 242 CCGAGTGAACCTTTGTCGAGAGAGAGACGCGCGCTTCATCGACGAGGCTAAGCGCG 301
DB 911 GCAAGTGAATGCTTACAGAGACCGAAAGCGCCCATTTCTTCGAGAGGACAGAGCG 970
OY 302 TGCGAGCGCTGCACATGAAAGAGACCCCGATTAATAATACCGCGCGCGGAGAAACCA 361
DB 971 TACAGCGGCTGCAGGAGATTAATACCGGCTATTAATACGACCTCGCGCAAGGAG 1030
OY 362 AGAGC 366
DB 1031 AGAGG 1035

RESULT 5
US-08-860-635A-18
Sequence 18, Application US/08860635A
Patent No. 6143878

GENERAL INFORMATION:
APPLICANT: Koopman, Peter
APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860, 635A
FILING DATE: 29-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3243 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-468-577B-6

```

```

Query Match          20.0%; Score 99.8; DB 3; Length 3243;
Best Local Similarity 62.8%; Pred. No. 1e-13;
Matches 155; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```

```

QY 129 CCGCCGACAGCCGCGCATGATGCTTCATGTCGTCGCCGCGGCGCAGCAT 188
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1015 CCACATCAAGAGCCGCGCATGATGTCGTCGCCGCGGCGCAGCATGAT 1074
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 GCGCCGAGAGACCCGCGCATGATGTCGTCGCCGCGGCGCAGCATGAT 248
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1075 CCGTCAAGCCTCCGACATGTCGTCGCCGCGGCGCAGCATGATGAT 1134
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 GAACTTTTGTGCGAGAGGAGGCGGCGCTTCATGACGAGGCTTAAGCGCTGCGAGC 308
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1135 GAAGTCATGACCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1194
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 309 GCTGCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 368
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1195 GCAGGACCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1254
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 CATGAG 375
    || |||
DB 1255 CGTGAG 1261

```

```

RESULT 10
US-08-239-276-2
; Sequence 2, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

```

```

; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/239,276
; FILING DATE: 05-MAY-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5051 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-239-276-2

```

```

Query Match          17.9%; Score 89.4; DB 2; Length 5051;
Best Local Similarity 60.2%; Pred. No. 2e-11;
Matches 147; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

```

QY 118 AACAGCCGAGCCGCTCAAGCGGCGCATGATGCTTCATGTCGTCGCCGCGGCGAG 177
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1800 AGCAGCGAGCCGACATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1859
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 178 GCGCCGAGAGTGGCCGCGAGAGACCCGAGATGCAACGTCGAGATGCAAGCGGCTG 237
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1860 AGGAGAGAAATCTTCAGGCGCTTCCCGCATGATGATCAACATTAAGCAAAATCTTA 1919
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 238 GCGCCGAGTGGAAATCTTGTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 297
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1920 GGATCTCGCTGGAATTCATGTCGCAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1979
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 298 GCGCTGCGAGGCTGCACATGAGAGGAGCAGCCGAGATTAATACCGGCGCGGAGAA 357
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1980 CGGCTAAGCAAGATCCACTTAAGAGAGTACCCAACTATTAATACAAACCGGAGGAGAA 2039
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 358 ACCA 361
    ||
DB 2040 CNCA 2043

```

```

RESULT 11
US-08-468-579B-2
; Sequence 2, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York

```

```

:
: COUNTRY: USA
: ZIP: 10591-5144
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: System 7.5
: SOFTWARE: WordPerfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468,579B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 530
:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/239,276
: FILING DATE: 05-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/872,646
: FILING DATE: 08-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/715,181
: FILING DATE: 14-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/441,703
: FILING DATE: 04-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/312,543
: FILING DATE: 17-FEB-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Kurt G. Briscoe
: REGISTRATION NUMBER: 33,141
: REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (914) 332-1700
: TELEFAX: (914) 332-1844
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5051 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-468-579B-2
:
: Query Match 17.9%; Score 89.4; DB 2; Length 5051;
: Best Local Similarity 60.2%; Pred. No. 2e-11;
: Matches 147; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
:
: Oy 118 AACAGCCGCGACCGCGTCAAGCGGCCCATGATGCTTCATGGTGTGTCGCCGCGGCAG 177
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1800 AGCAGCGAGCCACACATTAAGCGACATGATTCATTCATTCATTCATTCATTCATTCAT 1859
:
: Oy 178 CGCGCGAAGATGCGCCAGAGAGACCCCAAGATGCACTCGAGATCAGACGCGCTG 237
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1860 AGGAGAAAAATCCCTTCAGGCGCTTCGCCGACATGCATTAATCAACATTAAGCAAAATCTTA 1919
:
: Oy 238 GCGCGCGAGTGGAAACTTTTGTGCGAGACGAGAGCGCGCGTTCATCGACGAGCTAAG 297
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1920 GGAATTCGCTGGAATATCAATATCCACACGAGAGAGCAACCTTTATTAAGAGCAAGCC 1979
:
: Oy 298 CGCGTGCAGCGCTGCACATGAAAGAGACCCGAGTTATTAATACCGCGCCGCGGAGAA 357
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1980 CGGCTAAGCAAGATCCACTTAGAGAGATACCAACTATTAATACAAACCCGCGACGAAA 2039
:
: Oy 358 ACCA 361
: | |
: Db 2040 CNCA 2043
:
: RESULT 12
: US-08-468-577B-2
: Sequence 2, Application US/08468577B
: Patent No. 6001804
: GENERAL INFORMATION:
: APPLICANT: Rabin, Daniel
: TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

```

```

:
: TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
:
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sprung Kramer Schaefer & Briscoe
: STREET: 660 White Plains Road
: CITY: Tarrytown
: STATE: New York
: COUNTRY: USA
: ZIP: 10591-5144
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: System 7.5
: SOFTWARE: WordPerfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468,577B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 514
:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/239,276
: FILING DATE: 05-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/872,646
: FILING DATE: 08-JUN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/715,181
: FILING DATE: 14-JUN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/441,703
: FILING DATE: 04-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/312,543
: FILING DATE: 17-FEB-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Kurt G. Briscoe
: REGISTRATION NUMBER: 33,141
: REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (914) 332-1700
: TELEFAX: (914) 332-1844
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5051 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-468-577B-2
:
: Query Match 17.9%; Score 89.4; DB 3; Length 5051;
: Best Local Similarity 60.2%; Pred. No. 2e-11;
: Matches 147; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
:
: Oy 118 AACAGCCGCGACCGCGTCAAGCGGCCCATGATGCTTCATGGTGTGTCGCCGCGGCAG 177
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1800 AGCAGCGAGCCACACATTAAGCGACATGATTCATTCATTCATTCATTCATTCATTCAT 1859
:
: Oy 178 CGCGCGAAGATGCGCCAGAGAGACCCCAAGATGCACTCGAGATCAGACGCGCTG 237
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1860 AGGAGAAAAATCCCTTCAGGCGCTTCGCCGACATGCATTAATCAACCTTAAGCAAAATCTTA 1919
:
: Oy 238 GCGCGCGAGTGGAAACTTTTGTGCGAGACGAGAGCGCGCGTTCATCGACGAGCTAAG 297
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1920 GGAATTCGCTGGAATATCAATATCCACACGAGAGAGCAACCTTTATTAAGAGCAAGCC 1979
:
: Oy 298 CGCGTGCAGCGCTGCACATGAAAGAGACCCGAGTTATTAATACCGCGCCGCGGAGAA 357
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Db 1980 CGGCTAAGCAAGATCCACTTAGAGAGATACCAACTATTAATACAAACCCGCGACGAAA 2039
:
: Oy 358 ACCA 361
: | |
: Db 2040 CNCA 2043

```

RESULT 13

US-08-239-276-5
Sequence 5, Application US/08239276
Patent No. 5840836
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kiemer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/239, 276
FILING DATE: 05-MAY-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1570 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-239-276-5

Query Match 16.2%; Score 80.8; DB 2; Length 1570;
Best Local Similarity 62.9%; Pred. No. 1.3e-09;

Matches 141; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 129 CCGCGTCAGAGCGCCCAAGATGCTTCATGGTGTGTCGCCCGGGGACAGCGCGCAAGAT 188
DB 163 CCACATCAAGAGAGCCCATGACGCTTCATGGTGTGCGCAAGATGACGAGGAAGAT 222
QY 189 GGGCCAGAGAACCCCAAGATGACATCTCGAGATCAGCAAGCGCTTGGGGCCGAGTG 248
DB 223 CCGGAAGCGCTTCCAGACATGCAACATCCGACATCAGACAGATCTTGATCTGCTG 282
QY 249 GAACTTTTGTGCGAGAGAGAGAGCGCGCTTCATCGACGAGGCTAAGCGGCTGCGAGC 308
DB 283 GAAGTCATGACCAACAGAGAGAGAGCGCGCTTCATCGACGAGGCTTCTGATCG- 341
QY 309 GCTGCACATGAGAGAGACCGCGGATTAATAATACCGCGCCGCGC 352
DB 342 TCATCACTGAGAGAGATCTCTGACTACAGATACAGCGCGCGC 385

RESULT 14

US-08-468-579B-5
Sequence 5, Application US/08468579B
Patent No. 5981700
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kiemer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1570 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-579B-5

Query Match 16.2%; Score 80.8; DB 2; Length 1570;
Best Local Similarity 62.9%; Pred. No. 1.3e-09;

Matches 141; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 129 CCGCGTCAGAGCGCCCAAGATGCTTCATGGTGTGTCGCCCGGGGACAGCGCGCAAGAT 188
DB 163 CCACATCAAGAGAGCCCATGACGCTTCATGGTGTGCGCAAGATGACGAGGAAGAT 222
QY 189 GGGCCAGAGAACCCCAAGATGACATCTCGAGATCAGCAAGCGCTTGGGGCCGAGTG 248
DB 223 CCGGAAGCGCTTCCAGACATGCAACATCCGACATCAGACAGATCTTGATCTGCTG 282
QY 249 GAACTTTTGTGCGAGAGAGAGAGCGCGCTTCATCGACGAGGCTAAGCGGCTGCGAGC 308
DB 283 GAAGTCATGACCAACAGAGAGAGAGCGCGCTTCATCGACGAGGCTTCTGATCG- 341
QY 309 GCTGCACATGAGAGAGACCGCGGATTAATAATACCGCGCCGCGC 352

Db 342 TCATCACCCTGGAGAGATATCTGACTACACAGTACAGCCGGGC 385

RESULT 15
US-08-468-577B-5

; Sequence 5, Application US/08468577B
; Patent No. 6001804

; GENERAL INFORMATION:

; APPLICANT: Rabin, Daniel

; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sprung Kramer Schaefer & Briscoe

; STREET: 660 White Plains Road

; CITY: Tarrytown

; STATE: New York

; COUNTRY: USA

; ZIP: 10591-5144

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: System 7.5

; SOFTWARE: WordPerfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/468,577B

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/239,276

; FILING DATE: 05-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/872,646

; FILING DATE: 08-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/715,181

; FILING DATE: 14-JUN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/441,703

; FILING DATE: 04-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/312,543

; FILING DATE: 17-FEB-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Kurt G. Briscoe

; REGISTRATION NUMBER: 33,141

; REFERENCE/DOCKET NUMBER: MDI 251.8-KGB

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (914) 332-1700

; TELEFAX: (914) 332-1844

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1570 nucleotides

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-468-577B-5

Query Match 16.2%; Score 80.8; DB 3; Length 1570;

Best Local Similarity 62.9%; Pred. No. 1.3e-09;

Matches 141; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

OY 129 CCGGCTCAAGCGGGCCGATGATGCTGCTCCCGGGCGAGCGCGCAAGAT 188

Db 163 CCACATCAAGAGCGCCATGAGCTTATGCTGCGCAAGATGAGCGAGAGAT 222

OY 189 GCGCCAGAGAACCCCAAGATGACACACTCGAGATCAGCAAGCGCTGGCGCGAGTG 248

Db 223 CCTGCAAGCTCTCCAGATGACCAACTCCAGATCAGCAAGATCCTTGATCTCGCTG 282

OY 249 GAACCTTTGTGCGAGAGAGCGCGCTTCATCGAGAGGCTAGAGCGCTGCGAGC 308

Db 283 GAAGTCATGACCAACAGAGAGACAGCTTATGAGGAGACAGGCTTCTGATCG- 341

OY 309 GCTGCACATGAGAGAGCACCCGGATTATTAATCCGGCCCCGGGC 352

Db 342 TCATCACCCTGGAGAGATATCTGACTACACAGTACAGCCGGGC 385

Search completed: October 10, 2002, 17:33:01
Job time : 31.7333 secs

R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldi, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.livnl.gov/bbrp/image/image.html
 Insert length: 911 Std Error: 0.00
 Seq primer: 400p from Gibco
 High quality sequence stop: 462.
 Location/Qualifiers

FEATURES

source

1..537
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2480730"
 /clone_lib="NCI-CGAP_GCC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
 from the normalized library NCI-CGAP GC4 was prepared, and
 ss circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldi."
 BASE COUNT 124 a 161 c 190 g 60 t 2 others
 ORIGIN

Query Match 99.6%; Score 498; DB 9; Length 537;
 Best Local Similarity 99.6%; Pred. No. 1.7e-80;
 Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 CACAGCGCCCGCATGTATCAATGATGAGAGCGAGCTGAAGCCCGCGCCGCGCA 60
 2 CACAGCGCCCGCATGTATCAATGATGAGAGCGAGCTGAAGCCCGCGCCGCGCA 61
 61 ACTTCGG 120
 62 ACTTCGG 121
 121 AGCCCGGACCGCGTCAAGCGGCCCATGAATGCCCTTCATGCTGTCGCCGCGCAGCGG 180
 122 AGCCCGGACCGCGTCAAGCGGCCCATGAATGCCCTTCATGCTGTCGCCGCGCAGCGG 181
 181 CGCAAGATGGCCCGAGAGAAACCCAGATGCCAATCGCAATCGCAAGCGCCCTGGGC 240
 182 CGCAAGATGGCCCGAGAGAAACCCAGATGCCAATCGCAATCGCAAGCGCCCTGGGC 241
 241 GCGGAGTGAACCTTTGTCGAGAGCGAGAAAGCGCGCTTCATCGACGAGCTGAAGCGG 300
 242 GCGGAGTGAACCTTTGTCGAGAGCGAGAAAGCGCGCTTCATCGACGAGCTGAAGCGG 301
 301 CTGCGAGCGCTGACATGAAGAGCAACCGGATTAATAATACCGCGCCGCGGAGAAAC 360
 302 CTGCGAGCGCTGACATGAAGAGCAACCGGATTAATAATACCGCGCCGCGGAGAAAC 361
 361 AAGACGCTCATGAAGAAGATTAAGTACAGCTGCCGCGGCGCTGCTGGCCCGCGCGGC 420
 362 AAGACGCTCATGAAGAAGATTAAGTACAGCTGCCGCGGCGCTGCTGGCCCGCGCGGC 421
 421 AATAGCATGGCGGAGGG 480
 422 AATAGCATGGCGGAGGG 481
 481 ATGGACAGTTACGCGCACAT 500
 482 ATGGACAGTTACGCGCACAT 501

RESULT 7

BE206866 641 bp mRNA linear EST 27-JUN-2000
 LOCUS BE206866
 DEFINITION ba04909.y1 NIH-MGC.7 Homo sapiens cDNA clone IMAGE:2823424 5',
 similar to SW:SOX2_HUMAN P48431 TRANSCRIPTION FACTOR SOX-2. [1] ;,
 mRNA sequence.
 ACCESSION BE206866
 VERSION BE206866.1 GI:8750264
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 641)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library
 Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/ptml/lifresources.shtml
 Seq primer: 400p from Gibco
 High quality sequence stop: 518.
 Location/Qualifiers

FEATURES

source

1..641
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2823424"
 /clone_lib="NIH-MGC.7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pOTB1; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 136 a 205 c 231 g 69 t
 ORIGIN

Query Match 99.4%; Score 496.8; DB 9; Length 641;
 Best Local Similarity 99.6%; Pred. No. 2.9e-80;
 Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 CACAGCGCCCGCATGTATCAATGATGAGAGCGAGCTGAAGCCCGCGCCGCGCA 60
 63 CACAGCGCCCGCATGTATCAATGATGAGAGCGAGCTGAAGCCCGCGCCGCGCA 122
 61 ACTTCGG 120
 123 ACTTCGG 182
 121 AGCCCGGACCGCGTCAAGCGGCCCATGAATGCCCTTCATGCTGTCGCCGCGCAGCGG 180
 183 AGCCCGGACCGCGTCAAGCGGCCCATGAATGCCCTTCATGCTGTCGCCGCGCAGCGG 242
 181 CGCAAGATGGCCCGAGAGAAACCCAGATGCCAATCGCAATCGCAAGCGCCCTGGGC 240
 243 CGCAAGATGGCCCGAGAGAAACCCAGATGCCAATCGCAATCGCAAGCGCCCTGGGC 302
 241 GCGGAGTGAACCTTTGTCGAGAGCGAGAAAGCGCGCTTCATCGACGAGCTGAAGCGG 300
 303 GCGGAGTGAACCTTTGTCGAGAGCGAGAAAGCGCGCTTCATCGACGAGCTGAAGCGG 362
 301 CTGCGAGCGCTGACATGAAGAGCAACCGGATTAATAATACCGCGCCGCGGAGAAAC 360
 363 CTGCGAGCGCTGACATGAAGAGCAACCGGATTAATAATACCGCGCCGCGGAGAAAC 422

Qy	351	AAGAGCGTCATGAAGAAAGATAAATACACGCTGCCCGGCGGAGGCTGCTGGCCGCCCGGGGCG	420
Qy	352	AA	421
Qy	353	AA	422
Db	423	AAGAGCGCTCATGAAGAAAGATATACACGCTGCCCGGCGGAGGCTGCTGGCCGCCCGGGGCG	482
Qy	421	AATAGCATGGCGAGCGGGGCTGGGGGTGGCGCGCGGCTGGGGCTGGGGCTGGAAACACACGC	480
Qy	422	AA	481
Db	483	AATAGCATGGCGAGCGGGGTGGGGGTGGCGCGCGGCTGGGGCTGGGGCTGGAAACACACGC	542
Qy	481	AA	543
Qy	482	ATGAGCATATACGGGCACAT	500
Qy	483	AA	501
Db	543	ATGGACAGTACGGCCACAT	562

RESULT 8	LOCUS	DEFINITION	LOCUS	DEFINITION
BE672418	BE672418	728 bp mRNA	1 linear EST	08-SEP-2000
	7a60e05.x1	NCI-CGAP GC6 Homo sapiens cDNA clone IMAGE:322316 3'		
		similar to SW:SOX2_HUMAN P48411 TRANSCRIPTION FACTOR SOX-2. [1]		
		contains TAR1.tl TAR1 repetitive element ; mRNA sequence.		
		BE672418		

ACCESSION	BE672418	GI:10032959
VERSION	BE672418.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Eukariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997)
			Contact: Robert Strausberg, Ph.D.	

Email: cgapdb-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LML, send email to: info@image.lml.gov

FEATURES	
source	
	Seq primer: -40UP from Gibco
	High quality sequence stop: 497.
	Location/Qualifiers
	1. .728

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3223136"
/clone_lib="NCL-CGAP_GCC"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notice="Vector: pT73-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; Plasmid DNA
from the normalized library NCL-CGAP_GC4 was prepared, and
ss circles were made in vitro. FOLLOWING HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonesids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

```

Query Match	99.4%	Score 496.8	DB: 10	Length 728
Best Local Similarity	99.6%	Pred. No. 2.9e-80		
Matches 498	Conservative 0	Mismatches 2	Indels 0	Gaps 0

QY	1	CACACGCCCGCATGTATACATCATGTATGGAGACGAGACTTGAACGCCGCCGCCGACGACAA	60
Db	27	CACACGCCCGCATGTATACATCATGTATGGAGACGAGACTTGAACGCCGCCGCCGACGACAA	86

Oy	61	ACTTGGGGGGGGGGGGGCAACTCCACCGGGGGGGCGCGGCACCAAGAAAAAC	120
Dd	87	ACTTGGGGGGGGGGGGGGCACTCCACCGGGGGGGCGCGGCACCAAGAAAAAC	146
Oy	121	AGCCCGGACCGCGTCAGCGGGCCATGAATGCTTCATGTTGGTCCCGGGGACGCGG	180
Dd	147	AGCCCGGACCGCGTCACCGGGCCCATGTGATGCTTCATGTTGGTCCCGGGGACGCGG	206
Oy	181	CGCAGATGGCCCGAGAGACCCCAAGATGCGCACTGGAGATGATAGCAAGCGCCTGGGC	240
Dd	207	CGCAGATGGCCCGAGAGACCCCAAGATGCGCACTGGAGATGATAGCAAGCGCCTGGGC	266
Oy	241	GCCGAGTGAATCTTTTGTGCGAGACGGAGAGAGCGGCGTTTCATCGACAGGCTAAAGCG	300
Dd	267	GCCGAGTGAATCTTTTGTGCGAGACGGAGAGAGCGGCGTTTCATCGACAGGCTAAAGCG	326
Oy	301	CTGCGAGGCGTGCACATGAAGAGACCCCGGATTTTAAATACCGGGCCCCGGGGAAAAAC	360
Dd	327	CTGCGAGGCGTGCACATGAAGAGACCCCGGATTTTAAATACCGGGCCCCGGGGAAAAAC	386
Oy	361	AAGAGGTCATGAAGAAAGATTAATGACACGCGCCGGCGGGCTCTGAGCCCCGGCGAC	420
Dd	387	AAGAGGTCATGAAGAAAGATTAATGACACGCGCTGCGCCGGGGCTCTGAGCCCCGGCGAC	446
Oy	421	AATAGCATGGCGAGCGGGGTCGGGGTGGAGCGCGGCGCTGGCGCGCGCTGAAACCAAGCGC	480
Dd	447	AATAGCATGGCGAGCGGGTGTGGGGTGGAGCGCGGCGCTGGCGCGCGCTGAAACCAAGCGC	506
Oy	481	ATGGAAGTAAAGCGCACAT 500	
Dd	507	ATGGAAGTAAAGCGCACAT 526	

RESULT	9
A1571299	
LOCUS	507 bp mRNA linear EST 12-MAY-1999
DEFINITION	tta4e04.x1 NCI-CGAP_Brr25 Homo sapiens cDNA IMAGE:2170494 3'

similar to gb:U1136 SOX-10 PROTEIN (HUMAN):contains TAR1.t2 TAR1
repetitive element ;, mRNA sequence.
AI571299
AI571299.1 GI:4534673
EST.

SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 507)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI/NINDS-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTCAP), Tumor Gene Index

JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-1@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.

Ph.D.,
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Dantas, Ph.D.

BONAIACI, PH.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
RNA Sequencing by: Washington University Genome Sequencing Center

found through the T.M.A.G.E. Consortium/NIH, at:
Clone distribution: NCI-CGAP clone distribution information can be
sequencing center
Washington University genome sequencing center
sequencing of: Washington University genome sequencing center

www-bio.1nl.gov/bbrp/image/image.html
Insert length: 1184 Std Error: 0.00

Seq primer: -40UP from Gibco
High quality sequence stop: 451

[illegible]

```
source      1. .507
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"  
/clone="IMAGE:2170494"
```

```
/clone_lib="NCI_CGAP_Brn25"
```


Db	481	ATGGACACTGACGGCACAT	500
RESULT	11		
LOCUS	A1356682		
DEFINITION	A1356682	965 bp	mRNA
ACCESSION	q922a10.x1	NCI_CGAP_Brn23	Homo sapiens cDNA clone IMAGE:2012730 3'
VERSION	A1356682	similar to gb:X71136	SOX-10 PROTEIN (HUMAN);, mRNA sequence.
KEYWORDS	A1356682.1	GI:4108303	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
JOURNAL	1 (bases 1 to 965)		
COMMENT	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
	National Cancer Institute / National Institute of Neurological		
	Disorders and Stroke, Brain Tumor Genome Anatomy Project		
	(CGAP/BRAP), Tumor Gene Index		
	Unpublished (1998)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-rt@mail.nih.gov		
	Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,		
	Ph.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima		
	Bonaldo, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	www-bio.lnl.gov/bdrrp/image/image.html		
	Insert Length: 1190 Std Error: 0.00		
	Seq primer: -40UP from Gidco		
	High quality sequence stop: 480.		
FEATURES	Location/Qualifiers		
SOURCE	1..965		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2012730"		
	/clone_1lb="NCI_CGAP_Brn23"		
	/tissue_type="gliblastoma (pooled)"		
	/lab_host="DH10B"		
	/note="Organ: brain; Vector: pTV3D-Pac (Pharmacia) with a		
	modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer [5'		
	TGTATCCATCTGTAAGTGGAGCGCGCGGCAATCTTTTCTTTTCTTTTCTTTT		
	T 3']; double-stranded cDNA was ligated to Eco RI		
	adaptors (Pharmacia), digested with Not I and cloned into		
	the Not I and Eco RI sites of the modified pTV3 vector.		
	Library is normalized, and was constructed by Bento		
	Soares and M.Fatima Bonaldo."		
BASE COUNT	202 a 329 c 308 g 125 t	1	others
ORIGIN			
Query Match	97.6%;	Score 488;	DB 9;
Best Local Similarity	99.8%;	Pred. No. 1.2e-78;	
Matches 499;	Conservative 0;	Mismatches 0;	Indels 1;
			Gaps 1;
1	CACACGCGCCGATGTACACATGATGAGAGAGAGCGTGAAGCCGCGGCGCCGACGCA	60	
27	CACACGCGCCGATGTACACATGATGAGAGAGCGTGAAGCCGCGGCGCCGACGCA	86	
61	ACTTGG	120	
87	ACTTGG	146	
121	AGCCCGGAGCGCGGTCAAGCGGGCCCATGTGATGCTTCATGATGTGTGTCCCGCGGAGCGG	180	
147	AGCCCGGAGCGCGGTCAAGCGGGCCCATGTGATGCTTCATGATGTGTGTCCCGCGGAGCGG	206	
161	CGCAGATGGCCCGAGAGACCCCAAGATGCACCACTGGAGATAGCAAGCGCTGGGC	240	

Db	207	CGCAAGATGGGCCAGAGAAACCCCAAGATGCACAACTCGAGATCAGCAACGCCCTGGGC	266
QY	241	GGCGAGTGGAAACCTTTGTGGGAGCGGAGAGAGGGCGCTCATCGAGAGCTTAAGCG	300
Db	267	GCCGAGTGGAAACTTTTGTGGAGACGAGAGAGAGCGCGCTTCATCGACGAGGCTTAAGGG	326
QY	301	CTGCGAGCGCTGCACATGAGAGAGACCCCGGATTAATAATACCGGCCCCCGGGAAC	360
Db	327	CTGGAGAGCGCTGCACATGAGAGAGACCCCGGATTAATAATACCGGCCCCCGGGAAC	386
QY	361	AAGAGCGCTCATGAGAGAGATTAAGTACACGCTGCCCGGCGGCTGTGCCCGCGCGC	420
Db	387	AAGAGCGCTCATGAGAGAGATTAAGTACACGCTGCCCGGCGGCTGTGTGCCCGCGCGC	445
QY	421	AATAGACATGCGAGCGGGGTGGGGTGGGGCGCGGCTGTGGGCGCGGGCGTAAACAGCGC	480
Db	446	AATAGACATGCGAGCGGGGTGGGGTGGGGCGCGGCTGTGGGCGCGGGCGTAAACAGCGC	505
QY	481	ATGACACGTTACGGCGACACAT 500	
Db	506	ATGACACGTTACGGCGACACAT 525	
RESULT 12			
AW055151		607 bp	linear
LOCUS			EST 09-MAR-2000
DEFINITION			similar to gb:U1136 SOX-10 PROTEIN (HUMAN);, mRNA sequence.
ACCESSION			AW055151
VERSION			AW055151.1
KEYWORDS			EST.
SOURCE			human.
ORGANISM			Homo sapiens
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE			NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL			National Cancer Institute / National Institute of Neurological
COMMENT			Disorders and Stroke, Brain Tumor Genome Anatomy Project
			(CGAP/BTCAP), Tumor Gene Index
			Unpublished (1998)
			Contact: Robert Strausberg, Ph.D.
			Email: cgapsb-remail.nih.gov
			Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
			Ph.D.
			CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
			Bonaldo, Ph.D.
			CDNA Library Arrayed by: Greg Lennon, Ph.D.
			DNA Sequencing by: Washington University Genome Sequencing Center
			Clone distribution: NCI-CGAP clone distribution information can be
			found through the I.M.A.G.E. Consortium/LNL at:
			www-bio.llnl.gov/bdrrp/image/image.html
			Insert Length: 1252 Std Error: 0.00
			Seq primer: -40UP from Gldco
			High quality sequence stop: 501.
FEATURES			
SOURCE			
			1..607
			/organism="Homo sapiens"
			/db_xref="taxon:9606"
			/clone="IMAGE:2556783"
			/clone_lib="NCI CGAP Brn23"
			/tissue_type="glioblastoma (pooled)"
			/lab_host="DH10B"
			/note="Organ: brain; Vector: pUT3D-Pac (Pharmacia) with a
			modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
			strand cDNA was primed with a Not I...oligo(dT) primer; [5'
			TGGTCCAAATCTGAAGTGGAGCGCGCGCATATCTTTTATTTTATTTTATTTT
			T 3']; double-stranded cDNA was ligated to Eco RI
			adaptors (Pharmacia), digested with Not I and cloned into
			the Not I and Eco RI sites of the modified pUT3D vector.
			Library is normalized, and was constructed by Bento
			Soares and M. Fatima Bonaldo."
BASE COUNT			
			133 a 192 c 214 g 64 t 4 others

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
 Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 1142 Std Error: 0.00
 Seq primer: -40UP from G1bco
 High quality sequence stop: 485
 POLYA-No.

FEATURES Location/Qualifiers
 source 1..514
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2214163"
 /clone_1b="NCI-CGAP.Lu19"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 110 a 160 c 190 g 54 t
ORIGIN

Query Match 95.4%; Score 477; DB 9; Length 514;
 Best Local Similarity 99.8%; Pred. No. 1e-76;
 Matches 488; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CACAGCGCCCGCATGTACACATGATGAGAGCGAGCTGAAGCCCGCGCCGACGACAA 60
 |||||||
 DB 27 CACAGCGCCCGCATGTACACATGATGAGAGCGAGCTGAAGCCCGCGCCGACGACAA 86
 QY 61 ACTTCGGGGGGCGCGCGCACTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
 |||||||
 DB 87 ACTTCGGGGGGCGCGCGCGCACTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 146
 QY 121 AGCCCGAGCGCGTCAAGCGCGCCCATGATGCTTATGCTGTGTCGCCGGGGACGCG 180
 |||||||
 DB 147 AGCCCGAGCGCGTCAAGCGCGCCCATGATGCTTATGCTGTGTCGCCGGGGACGCG 206
 QY 181 CGCAAGATGGCCAGAGAGAACCCCAAGATGACAACTCGAGATCGAGAAAGCCCTGGCG 240
 DB 207 CGCAAGATGGCCAGAGAGAACCCCAAGATGACAACTCGAGATCGAGAAAGCCCTGGCG 266
 QY 241 GCCGAGTGAACCTTTTGTTCGAGAGAGCGAGAAAGCGCGCTTCATCGACGAGCTTAAGCG 300
 |||||||
 DB 267 GCCGAGTGAACCTTTTGTTCGAGAGAGCGAGAAAGCGCGCTTCATCGACGAGCTTAAGCG 326
 QY 301 CTCGAGCGCTGCACATGAAGAGACCGCGGATTAATATACCGCGCCCGCGGAGAAAC 360
 |||||||
 DB 327 CTCGAGCGCTGCACATGAAGAGACCGCGGATTAATATACCGCGCCCGCGGAGAAAC 386
 QY 361 AAGAGCGCTATGAAGAGATAGTACAGCGCTGCCCGCGCGCGCTGCTGGCCCGCGCGCGC 420
 |||||||
 DB 387 AAGAGCGCTATGAAGAGATAGTACAGCGCTGCCCGCGCGCGCTGCTGGCCCGCGCGCGC 446

QY 421 AATAGCATGGCGAGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
 |||||||
 DB 447 AATAGCATGGCGAGCGCGGTC-66GTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 505
 QY 481 ATGACAGT 489
 |||||||
 DB 506 ATGACAGT 514

RESULT 15
LOCUS BE792833 740 bp mRNA linear EST 20-SEP-2000
DEFINITION 601584834F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3939033 5',
 mRNA sequence.
ACCESSION BE792833
VERSION BE792833.1 GI:10214031
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 740)
AUTHORS NIH-MGC <http://imgc.ncbi.nlm.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L10M787 row: d column: 10
 High quality sequence stop: 740.

FEATURES Location/Qualifiers
 source 1..740
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3939033"
 /clone_1b="NIH_MGC_7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pOD5; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(C). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 158 a 240 c 249 g 93 t
ORIGIN

Query Match 93.4%; Score 467; DB 10; Length 740;
 Best Local Similarity 100.0%; Pred. No. 6.8e-75;
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GAGCTGAAGCG 93
 |||||||
 DB 1 GAGCTGAAGCG 60
 QY 94 GCGGCG 153
 |||||||
 DB 61 GCGGCG 120
 QY 154 TTCATGTGTGTCGCCCG 213
 |||||||
 DB 121 TTCATGTGTGTCGCCCG 180
 QY 214 AACTCGAGATGAGCAAGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 273
 |||||||

```

Db 181 AACTGGAGATCAGCAAGCCCTGGGCGCCGAGTGAACCTTTGTGAGAGACGAGAAAG 240
QY 274 CGGCCGTTTCATCGACGAGGCTTAAGCGGCTGCGAGCGCTGCACATGAAGAGCACCCGAT 333
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CGGCCGTTTCATCGACGAGGCTTAAGCGGCTGCGAGCGCTGCACATGAAGAGCACCCGAT 300
QY 334 TATTAATACCGGCGCCCGGCGGAAACCAAGACGCTTCATGAAGAAGGATAAGTACAGCTG 393
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 TATTAATACCGGCGCCCGGCGGAAACCAAGACGCTTCATGAAGAAGGATAAGTACAGCTG 360
QY 394 CGGCCGAGGCTGCTGGCCCCCGGCGGCAATAGCATGCGAGCGGGTGGGGTGGGGGCC 453
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 CGGCCGAGGCTGCTGGCCCCCGGCGGCAATAGCATGCGAGCGGGTGGGGTGGGGGCC 420
QY 454 GGCTGGGGCGGCGGCTGAACCAAGGCAATGACAGTTACGGGCACAT 500
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GGCTGGGGCGGCGGCTGAACCAAGGCAATGACAGTTACGGGCACAT 467

```

Search completed: October 10, 2002, 20:27:13
 Job time : 702.733 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 : Search time 84.9333 Seconds
(without alignments)
10107.424 Million cell updates/sec

Title: US-09-489-101a-3_COPY_1_500
Perfect score: 500
Sequence: 1 cacagcgcgcgcgtacacaa.....atgacagcttacgcgcacat 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : N_Geneseq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	1085	22	AAD11110 Human small cell 1
2	500	100.0	1087	22	AAK52388 Human polynucleoti
3	500	100.0	1116	22	AAK53268 Human polynucleoti
4	500	100.0	1126	22	AAK52284 Human polynucleoti
5	500	100.0	1501	22	AAK53372 Human polynucleoti
6	432.8	86.6	2283	24	AB199318 Mouse ischaemic co
7	350	70.0	395	24	AAK51666 Lung small cell ca
8	279	55.8	1542	20	AAK16151 Human Sox1 encodin
9	279	55.8	4091	22	AAD11111 Human small cell 1

10	266.2	53.2	2312	20	AAK16152 Chicken Sox1 cDNA.
11	261.4	52.0	2376	20	AAK16153 Mouse Sox1 cDNA.
12	234.8	47.0	2309	22	AAD11118 Human small cell 1
13	225	45.0	8372	22	AAD11119 Human small cell 1
14	220	44.0	2378	23	AAK87738 DNA encoding novel
15	217.8	43.6	412	24	AAK51625 Lung small cell ca
16	209.2	41.8	1161	20	AAK16154 Degenerate human S
17	199	39.8	1969	23	AAK87779 DNA encoding novel
18	193.4	38.7	2298	23	AB118481 Drosophila melanog
19	193.4	38.7	4298	23	AB118480 Drosophila melanog
20	150.8	30.2	1788	22	AAK16401 Human cDNA sequenc
21	143.2	28.6	3934	17	AAK130309 Human SOX-9 cDNA.
22	140.8	28.2	434	23	AAK87736 DNA encoding novel
23	140	28.0	2814	22	AAK22715 Human gastric canc
24	140	28.0	2514	17	AAK130308 Mouse Sox-9 cDNA.
25	138.4	27.7	2514	13	AAK20684 Bovine sex-determ
26	134.6	26.9	1661	16	AAK24814 Bovine sex-determ
27	134.6	26.9	1661	16	AAK24811 Bovine sex-determ
28	134.6	26.9	1661	16	AAK24813 Sex-determining re
29	134.6	26.9	2840	17	AAK13008 Bovine SRY-related
30	134.2	26.8	996	13	AAK20680 Rabbit mt-box. Or
31	128.6	25.7	1422	18	AAK159844 SOX4 coding sequen
32	128.2	25.6	1667	16	AAK20685 Porcine sex-determ
33	127.4	25.5	2069	24	AB199491 Mouse ischaemic co
34	121.4	24.3	441	24	AAK51805 Lung small cell ca
35	120.6	24.1	2033	21	AAK61724 Clone WGT24 of a g
36	119	23.8	2215	24	AB199688 Mouse ischaemic co
37	117.6	23.5	2310	23	AB105611 Drosophila melanog
38	117.4	23.5	10266	17	AAK13007 Mouse mt-box. Mus
39	117.4	23.5	10266	17	AAK13007 Mouse SRY-related
40	117.4	23.5	14704	13	AAK20685 PK5 741 insert con
41	117.2	23.4	4467	22	AAK2650 Human cervical can
42	115.2	23.0	7108	23	AB105610 Drosophila melanog
43	112.6	22.5	2762	16	AAK27042 Fish protamine gen
44	106.2	21.2	1588	18	AAK16167 Mouse Sox-18 gene.
45	103.8	20.8	317	16	AAK03988 Bovine testicle de

ALIGNMENTS

RESULT 1	
AAD11110	
ID	AAD11110 standard; DNA; 1085 BP.
XX	
AC	AAD11110;
XX	
DT	24-SEP-2001 (first entry)
XX	
AC	
DE	Human small cell lung cancer associated gene, SOX2.
XX	
XX	Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;
KW	melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW	leiomyosarcoma; synovial sarcoma; cytostatic; SOX2; ds.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	CDS
FT	/*tag= a
FT	/product= "Human SOX2 protein"
XX	
PN	MO20015349-A2.
XX	
PD	26-JUL-2001.
XX	
PF	19-JAN-2001; 2001WO-US02015.
XX	
PR	21-JAN-2000; 2000US-0489101.
XX	
PA	(LUDW-) LUDWIG INST. CANCER RES.
PA	(SLOK) SLOAN KETTERING INST. CANCER RES.
PA	(CORR) CORNELL RES. FOUND INC.

Db 154 AGCCGAGCCGCTCAAGCCGCCCATGATGCTTTCATGTTGTGTCCTCCGGGGACGCG 213
 Oy 181 CGAAGATGCGCCAGAGAGACCCCAAGATGACAACTCGGAGATCAGAACGCGCTGGGC 240
 Db 214 CGCAAGATGCGCCAGAGAGACCCCAAGATGACAACTCGGAGATCAGAACGCGCTGGGC 273
 Oy 241 GCCGAGTGAAGCTTTTGTGAGAGAGAGAGCGGCTTCATCGAGAGAGCTTAAGCGG 300
 Db 274 GCCGAGTGAAGCTTTTGTGAGAGAGAGAGCGGCTTCATCGAGAGAGCTTAAGCGG 333
 Oy 301 CTGCGAGCGCTGACATGAG 360
 Db 334 CTGCGAGCGCTGACATGAG 393
 Oy 361 AAGAGCGCTCATGAG 420
 Db 394 AAGAGCGCTCATGAG 453
 Oy 421 AATAGCATGCGAGAGCGGGGTGCGGGTGGCGCCGCGCTGGCGCGCGGCTGAACGCGC 480
 Db 454 AATAGCATGCGAGAGCGGGGTGCGGGTGGCGCCGCGCTGGCGCGCGGCTGAACGCGC 513
 Oy 481 ATGACAGTTACGCGCACAT 500
 Db 514 ATGACAGTTACGCGCACAT 533

RESULT 3

ID AAK53268 standard; cDNA; 1116 BP.
 AAK53268;

Dt 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 2797.
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW tissue growth factor; stem cell growth factor; haematopoiesis;
 KW nervous system disorder; arthritis; inflammation; ss.
 OS Homo sapiens.
 PN WO200157190-A2.
 PD 09-AUG-2001.
 PF 05-FEB-2001; 2001WO-US04098.
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 PA (HYSE-) HYSEQ INC.
 PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR P-PSDB; AAM80135.
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 PS Claim 1; Page 4995; 6221pp; English.
 XX

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation and which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibit activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX

SO Sequence 1116 BP; 256 A; 362 C; 362 G; 136 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 1116;
 Best Local Similarity 100.0%; Pred. No. 7.7e-95;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CACAGCGCCGCGATGTAACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
 Db 27 CACAGCGCCGCGATGTAACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 86
 Oy 61 ACTTGG 120
 Db 87 ACTTGG 146
 Oy 121 AGCCGAGCGCGTCAAGCGCGGCCCATGATGCTTCATGCTGTGTGTCTCCGCGGCGAGCGG 180
 Db 147 AGCCGAGCGCGTCAAGCGCGGCCCATGATGCTTCATGCTGTGTGTGTCTCCGCGGCGAGCGG 206
 Oy 181 CGCAAGATGCGCCAG 240
 Db 207 CGCAAGATGCGCCAG 266
 Oy 241 GCCGAGTGAAGCTTTTGTGAG 300
 Db 267 GCCGAGTGAAGCTTTTGTGAG 326
 Oy 301 CTGCGAGCGCTGACATGAG 360
 Db 327 CTGCGAGCGCTGACATGAG 386
 Oy 361 AAGAGCGCTCATGAG 420
 Db 387 AAGAGCGCTCATGAG 446
 Oy 421 AATAGCATGCGAGAGCGGGGTGCGGGTGGCGCCGCGCTGGCGCGCGGCTGAACGCGC 480
 Db 447 AATAGCATGCGAGAGCGGGGTGCGGGTGGCGCCGCGCTGGCGCGCGGCTGAACGCGC 506
 Oy 481 ATGACAGTTACGCGCACAT 500
 Db 507 ATGACAGTTACGCGCACAT 526

RESULT 4

ID AAK52284 standard; cDNA; 1126 BP.
 AAK52284;

Dt 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 829.
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX

OS Homo sapiens.
 XX MO200157190-A2.
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001MO-US04098.
 XX
 XX 03-FEB-2000; 2000US-0496914.
 XX 27-APR-2000; 2000US-0560875.
 XX 20-JUN-2000; 2000US-0598075.
 XX 19-JUL-2000; 2000US-0620325.
 XX 01-SEP-2000; 2000US-0654936.
 XX 15-SEP-2000; 2000US-0663561.
 XX 20-OCT-2000; 2000US-0693325.
 XX 30-NOV-2000; 2000US-0728422.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 XX Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX
 XX WPI: 2001-476283/51.
 XX P-PSDB; AAM79151.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 XX useful in diagnosis and gene therapy -
 XX Claim 1; Page 2766-2767; 6221pp; English.
 XX
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 XX encoded polypeptides (AAM78323-AAK60302) that exhibit activity relating to
 XX cytokine, cell proliferation or cell differentiation or which may induce
 XX production of other cytokines in other cell populations. The
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
 XX peptide therapy. The polypeptides have various cytokine-like activities,
 XX e.g. stem cell growth factor activity, haematopoiesis regulating
 XX activity, tissue growth factor activity, immunomodulatory activity and
 XX activity/inhibin activity and may be useful in the diagnosis and/or
 XX treatment of cancer, leukemia, nervous system disorders, arthritis and
 XX inflammation.
 XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 XX (AAM6020) are omitted as the relevant pages from the sequence listing
 XX were missing at the time of publication.
 XX
 XX Sequence 1126 BP; 262 A; 361 C; 364 G; 139 T; 0 other;
 XX
 XX Query Match 100.0%; Score 500; DB 22; Length 1126;
 XX Best Local Similarity 100.0%; Pred. No. 7,7e-95;
 XX Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1 CACAGCGCGCGCATGTACACATATGAGAGAGCTGAAGCGCGCGCGCGAGCA 60
 XX DB 34 CACAGCGCGCGCATGTACACATATGAGAGAGCTGAAGCGCGCGCGCGAGCA 93
 XX QY 61 ACTTCGG 120
 XX DB 94 ACTTCGG 153
 XX QY 121 AGCCCGGAGCG 180
 XX DB 154 AGCCCGGAGCG 213
 XX QY 181 CGCAAGATGGCG 240
 XX DB 214 CGCAAGATGGCG 273
 XX QY 241 GCCGAGTGGAACTTTTGTGCGAGACGAGAGCGCGCGCGCGCGCGCGCGCG 300
 XX DB 274 GCCGAGTGGAACTTTTGTGCGAGACGAGAGCGCGCGCGCGCGCGCGCGCG 333
 XX QY 301 CTGGAGCGCTGCACATGAAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 360

DB 334 CTGGAGCGCTGCACATGAAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
 QY 361 AAGACGCTCATGAAGAAGATAGTACAGCTGCCCGCGCGCGCGCGCGCGCGCGCG 420
 DB 394 AAGACGCTCATGAAGAAGATAGTACAGCTGCCCGCGCGCGCGCGCGCGCGCGCG 453
 QY 421 AATGACATGGCGAGCGGGGTCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
 DB 454 AATGACATGGCGAGCGGGGTCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
 QY 481 ATGACAGTTACCGGCACAT 500
 DB 514 ATGACAGTTACCGGCACAT 533
 RESULT 5
 AAK53372
 ID AAK53372 standard; cDNA; 1501 BP.
 XX
 XX AAK53372;
 XX
 XX 06-NOV-2001 (first entry)
 XX
 XX Human polynucleotide SEQ ID NO 2901.
 XX
 XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 XX tissue growth factor; immunomodulatory; cancer; leukemia;
 XX nervous system disorder; arthritis; inflammation; ss.
 XX
 XX Homo sapiens.
 XX
 XX MO200157190-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001MO-US04098.
 XX
 XX 03-FEB-2000; 2000US-0496914.
 XX 27-APR-2000; 2000US-0560875.
 XX 20-JUN-2000; 2000US-0598075.
 XX 19-JUL-2000; 2000US-0620325.
 XX 01-SEP-2000; 2000US-0654936.
 XX 15-SEP-2000; 2000US-0663561.
 XX 20-OCT-2000; 2000US-0693325.
 XX 30-NOV-2000; 2000US-0728422.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 XX Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX
 XX WPI: 2001-476283/51.
 XX P-PSDB; AAM60239.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 XX useful in diagnosis and gene therapy -
 XX Claim 1; Page 5070; 6221pp; English.
 XX
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 XX encoded polypeptides (AAM78323-AAK60302) that exhibit activity relating to
 XX cytokine, cell proliferation or cell differentiation or which may induce
 XX production of other cytokines in other cell populations. The
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
 XX peptide therapy. The polypeptides have various cytokine-like activities,
 XX e.g. stem cell growth factor activity, haematopoiesis regulating
 XX activity, tissue growth factor activity, immunomodulatory activity and
 XX activity/inhibin activity and may be useful in the diagnosis and/or
 XX treatment of cancer, leukemia, nervous system disorders, arthritis and
 XX inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAB60020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.

XX Sequence 1501 BP; 391 A; 428 C; 450 G; 220 T; 12 other;

Query Match 100.0%; Score 500; DB 22; Length 1501;
Best Local Similarity 100.0%; Pred. No. 7.8e-95;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CACAGCGCCCGCATGTACACATGATGAGACGGAGCTGAACCGCGCGCGCGAGCAA 60
   |||||||
DB 27 CACAGCGCCCGCATGTACACATGATGAGAGAGCTGAACCGCGCGCGCGAGCAA 86
   |||||||
DB 61 ACTTCGGGGGGGGGGGGGGGCAACTCCACCGCGGGCGCGCGCGCAACGAAAAAC 120
   |||||||
DB 87 ACTTCGGGGGGGGGGGGGGGCAACTCCACCGCGGGCGCGCGCGCAACGAAAAAC 146
   |||||||
QY 121 AGCCCGGACCGCGCTCAGCGCGCGCCATGAAATGCTTCATGTTGGTCCCGGGGACGG 180
   |||||||
DB 147 AGCCCGGACCGCGCTCAGCGCGCGCCATGAAATGCTTCATGTTGGTCCCGGGGACGG 206
   |||||||
QY 181 CGCAAGATGGCCGAGAGAACCCCAAGATGACACACTCGAGATCAGAAAGCTGAGG 240
   |||||||
DB 207 CGCAAGATGGCCGAGAGAACCCCAAGATGACACACTCGAGATCAGAAAGCTGAGG 266
   |||||||
QY 241 GCCGAGTGGAACTTTTGTGCGAGACGGAGAAAGCGCGCTTCATGACGAGGCTAAGCGG 300
   |||||||
DB 267 GCCGAGTGGAACTTTTGTGCGAGACGGAGAAAGCGCGCTTCATGACGAGGCTAAGCGG 326
   |||||||
QY 301 CTGCGAGCGCTGCACATGAAGAGACCGCGGATTAATAATACCGCGCGCGCGGAAAAAC 360
   |||||||
DB 327 CTGCGAGCGCTGCACATGAAGAGACCGCGGATTAATAATACCGCGCGCGCGGAAAAAC 386
   |||||||
QY 361 AAGAGCGTCATGAAGAGATGAATGATACAGCTGCCCGCGGGCTGTGTCGCCCGCGCGG 420
   |||||||
DB 387 AAGAGCGTCATGAAGAGATGAATGATACAGCTGCCCGCGGGCTGTGTCGCCCGCGCGG 446
   |||||||
QY 421 AATAGCATGGCGAGCGGGGTGCGGGTGGCGCGCGCTGGCGCGGGCGGTGAACCAAGCG 480
   |||||||
DB 447 AATAGCATGGCGAGCGGGGTGCGGGTGGCGCGCGCTGGCGCGGGCGGTGAACCAAGCG 506
   |||||||
QY 481 AATGACAGTTACGCGCACAT 500
   |||||||
DB 507 AATGACAGTTACGCGCACAT 526
   |||||||
```

RESULT 6
AB199318
ID AB199318 standard; cDNA; 2283 BP.

XX AB199318;

XX 07-MAR-2002 (first entry)

DE Mouse ischemic condition related cDNA sequence SEQ ID NO:179.

XX Mouse; ischemia; compressive ischemia; occlusive ischemia;
KW vasospastic ischemia; ischemic condition; ischemic disease; ss.

XX Mus musculus.

XX WO200188188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001MO-JP04192.

XX 18-MAY-2000; 2000JP-0145977.

XX (UYNT-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX WPI; 2002-034733/04.
DR P-PSDB; ABB57082.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -

PS Claim 2; Page 518-521; 2690pp; English.

XX The present invention describes a method for examining ischemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischemic condition (e.g. compressive
CC ischemia, occlusive ischemia or vasospastic ischemia) by measuring
CC expression levels of particular genes (AB199202 to AB199912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischemic condition-improving
CC drugs or therapeutics for ischemic diseases. AB199913 and AB199914
CC represent PCR primers for a mouse ischemic condition related sequence,
CC which are used in the exemplification of the present invention.

XX Sequence 2283 BP; 597 A; 583 C; 614 G; 488 T; 1 other;

Query Match 86.6%; Score 432.8; DB 24; Length 2283;
Best Local Similarity 92.5%; Pred. No. 6.6e-81;
Matches 468; Conservative 0; Mismatches 32; Indels 6; Gaps 1;

```
QY 1 CACAGCGCCCGCATGTACACATGATGAGACGGAGCTGAACCGCGCGCGCGAGCAA 60
   |||||||
DB 350 CCCAGCGCCCGCATGTATATACATGATGAGAGAGAGCTGAACCGCGCGCGCGAGCAA 409
   |||||||
QY 61 ACTTCGGGGGGGGGGGGGCGCGC-----GGCAACTCCACCGCGGGCGCGCGCGCAACACAG 114
   |||||||
DB 410 GCTTCGGGGGGGGGGGGGCGCGCGAGGAGGCAACCGCGGGCGCGCGCGCAACACAG 469
   |||||||
QY 115 AAAAAAGCGCGCGGACCGCTCAAGCGCGCGCGCATGAATGCTTCATGTTGGTGTCCCGGGG 174
   |||||||
DB 470 AAAAAAGCGCGCGGACCGCTCAAGCGCGCGCATGAATGCTTCATGTTGGTGTCCCGGGG 529
   |||||||
QY 175 CAGCGCGCGCAAGATGGCGCGAGAGAACCCCAAGATGCACAACTCGAGATCAGCAACGCG 234
   |||||||
DB 530 CAGCGCGCGTAAGATGGCGCGAGAGAACCCCAAGATGCACAACTCGAGATCAGCAACGCG 589
   |||||||
QY 235 CTGGGGCGCGAGTGAACCTTTTGTGCGAGACGGAGAAAGCGCGCTTCATCGACGAGGCT 294
   |||||||
DB 590 CTGGGGCGCGAGTGAACCTTTTGTGCGAGACGGAGAAAGCGCGCTTCATCGAGGAGGCG 649
   |||||||
QY 295 AAGCGGCTGCGAGCGCTGACATGAAGAGACCGCGGATTAATAATACCGCGCGCGCGG 354
   |||||||
DB 650 AAGCGGCTGCGGCTCTGCATGAAGAGACCGCGGATTAATAATACCGCGCGCGCGG 709
   |||||||
QY 355 AAAACCAAGACGCTCATGAAGAAGATGAATGATACAGCTGCCCGGGCGCTGCTGCCCGCC 414
   |||||||
DB 710 AAAACCAAGACGCTCATGAAGAAGATGAATGATACAGCTGCCCGGGCGCTGCTGCCCGCC 769
   |||||||
QY 415 GCGGCAATAGCATGGCGAGCGGGGTGCGGGTGGCGCGCGCGCGCGCGCGCGGCTGAC 474
   |||||||
DB 770 GCGGCAAGCATGGCGAGCGGGGTGCGGGTGGCGCGCGCGCGCGCGCGCGGCTGAC 829
   |||||||
QY 475 CAGCGCATGACAGTTACGCGCACAT 500
   |||||||
DB 830 CAGCGCATGACAGTTACGCGCACAT 855
   |||||||
```

RESULT 7
AAS61666
ID AAS61666 standard; cDNA; 395 BP.


```
XX Sequence 1542 BP; 238 A; 565 C; 560 G; 179 T; 0 other:
SQ
Query Match          55.8%; Score 279; DB 20; Length 1542;
Best Local Similarity 80.3%; Pred. No. 4,5e-49;
Matches 327; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 65 CGGGGGGCGGGCGGCACTCCACCGCGGGCGGGCGGCAACCAAGAAAACAGCC 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 142 CGGGGGGCGGGGGGGGAGAGCGGGGGCGGGCGGGGGGGGGCGCAAGGCCAACCC 201
QY 125 CGGACCGCGTCAACCGGCCCTGTAATGCTTCATGCTGTGTCGCCGGGCGGCGCA 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 202 AGGACCGGGTCAACCGGCCCTTCAATGCTGTGTCGCCGGGCGGCGCGCA 261
QY 185 AGATGGCCCAAGAGACACCCCAAGATGCACACTCGAGATCAGCAAGCGCTGGGCGCG 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 262 AGATGGCCCAAGAGACACCCCAAGATGCACACTCGAGATCAGCAAGCGCTGGGCGCG 321
QY 245 AGTGAACCTTTTGTCTCGGAGACGGAAGCGGCGCTTCATCGACGAGCTTAAGCGGCTGC 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 322 AGTGAAGCTATGTCTCGAGGCGCGGAGAGCGGCGCTTCATCGACGAGCGCAAGCGGCTGC 381
QY 305 GAGGCGTCCACATAGAGAGACCGCGGATTAATACCGCGCGCGGCGGCAAAACCAAGA 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 382 GCGGCTGCACATGAAGAGACACCGCGGATTAACAGTACCGGCGCGCGCAAGACCAAGA 441
QY 365 CGCTCATGAAGAGAGATTAAGTACACGCTGCCGCGGGGCTGCTGGCCCGCGGCAATA 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 442 CGCTGCTCAAGAGAGACCAAGTACTGCTGGCGGGGCTCTGCGGGCGCGGCGGCGGTG 501
QY 425 GCATGGCGAGCGGGGTGCGGGTGGGGCGCGGCTTGGCGCGGCGCGGTG 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 502 GCGGCGGCGCGGCTGTGGCCATGGGCGTGGGCGTGGCGCTGGGCGCGG 548

RESULT 9
AADI1111
ID AADI1111 standard; DNA; 4091 BP.
AC AADI1111;
XX
XX 24-SEP-2001 (first entry)
DE Human small cell lung cancer associated gene, SOX1.
XX
XX Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW leiomyosarcoma; synovial sarcoma; cytostatic; SOX1; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 61..1224
FT /*tag= a
FT /product= "Human SOX1 protein"
XX
XX WO200153349-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US02015.
XX
XX 21-JAN-2000; 2000US-0489101.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX (SLOK) SLOAN KETTERING INST CANCER RES.
XX (CORR) CORNELL RES FOUND INC.
XX
XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX WPI; 2001-457597/49.
XX
XX P-PSDB; AAE05811.
XX
```

```
XX Isolated polypeptide, used to treat or prognose a disorder
PT characterized by expression of a hCAAP e.g. cancer, is encoded by an
PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
XX
XX Claim 57; Page 88-90; 152pp; English.
XX
XX The invention relates to nucleic acids and encoded polypeptides which
CC are cancer associated antigens expressed in patients afflicted with
CC small cell lung cancer. The molecules provided by the invention can be
CC used in the diagnosis, monitoring, research or treatment of conditions
CC characterised by the expression of one or more cancer associated
CC antigens. The polypeptide is used to treat a disorder characterised by
CC expression of a hCAAP, and determine regression, progression or onset
CC of a condition characterised by expression of an abnormal amount of a
CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
CC are small and non-small cell lung cancer, melanoma, colon, breast, head
CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
CC The present sequence is a small cell lung cancer associated gene
CC designated as NT-SCIC-2, encoding human SOX1 protein.
XX
XX
SQ Sequence 4091 BP; 924 A; 1109 C; 1110 G; 943 T; 5 other:
Query Match          55.8%; Score 279; DB 22; Length 4091;
Best Local Similarity 80.3%; Pred. No. 4,7e-49;
Matches 327; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 65 CGGGGGGCGGGCGGCACTCCACCGCGGGCGGGCGGCGGCAACCAAGAAAACAGCC 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 143 CGGGCGGCGGGGGGGGAGAGCGGGGGCGGGCGGGGGGGCGCCCAAGGCCAACCC 202
QY 125 CGGACCGCGTCAACCGGCCCTGTAATGCTTCATGCTGTGTCGCCGGGCGGCGCA 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 203 AGGACCGGGTCAACCGGCCCTTCAATGCTGTGTCGCCGGGCGGCGCGCA 262
QY 185 AGATGGCCCAAGAGACACCCCAAGATGCACACTCGAGATCAGCAAGCGCTGGGCGCG 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 263 AGATGGCCCAAGAGACACCCCAAGATGCACACTCGAGATCAGCAAGCGCTGGGCGCG 322
QY 245 AGTGAACCTTTTGTCTCGGAGACGGAAGCGGCGCTTCATCGACGAGCTTAAGCGGCTGC 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 323 AGTGAAGCTATGTCTCGAGGCGCGGAGAGCGGCGCTTCATCGACGAGCGCAAGCGGCTGC 382
QY 305 GAGGCGTCCACATGAAGAGACACCGCGGATTAATACCGCGCGCGGCGGCAAAACCAAGA 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 383 GCGGCTGCACATGAAGAGACACCGCGGATTAACAGTACCGGCGCGCGCAAGACCAAGA 442
QY 365 CGCTCATGAAGAGAGATTAAGTACACGCTGCCGCGGGGCTGCTGGCCCGCGGCAATA 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 443 CGCTGCTCAAGAGAGACCAAGTACTGCTGGCGGGGCTCTGCGGGCGGCGGCGGTG 502
QY 425 GCATGGCGAGCGGGGTGCGGGTGGGGCGCGGCTTGGCGCGGCGCGGTG 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 503 GCGGCGGCGCGGCTGTGGCCATGGGCGTGGGCGTGGGCGGTG 549

RESULT 10
AAK16152
ID AAK16152 standard; CDNA to mRNA; 2312 BP.
AC AAK16152;
XX
XX 22-APR-1999 (first entry)
DE Chicken Sox1 cDNA.
XX
XX Chicken; Sox1; neuronal stem cell gene; neuroblastic cell; cancer;
XX nervous system; neurological disorder; ss.
XX
XX Gallus gallus.
XX
XX WO9900516-A2.
XX
XX
```

PD 07-JAN-1999.
 XX
 PF 25-JUN-1998; 98WO-GB01862.
 XX
 PR 25-JUN-1997; 97GB-0013469.
 XX
 PA (MED1-) MEDICAL RES COUNCIL.
 XX
 PI Lovell-Badge R, Pevny LH, Smith A;
 XX
 DR WPI; 1999-095759/08.
 XX
 PT Isolating neuroblastic cells from a population - by detecting the
 PT expression of the Sox1 gene in the cells and sorting the cells to
 PT isolate those cells expressing Sox1
 XX
 PS Disclosure; Page 41-43; 60pp; English.
 XX
 CC A method has been developed for isolating neuroblastic cells from a cell
 CC population. The method comprises: (a) detecting Sox1 gene expression in
 CC the cells; and (b) isolating those cells expressing Sox1. Also described
 CC is a method for producing a cell committed to the neuronal lineage,
 CC comprising: (a) transfecting a pluripotent stem cell with a genetic
 CC construct encoding Sox1 expression; (b) culturing the stem cells to
 CC differentiate into neural cells; and (c) isolating those neural cells
 CC produced. The present sequence represents chicken Sox1 cDNA. Detection
 CC of Sox1 expressing cells is important in diagnosing and treating cancers
 CC of the nervous system. Neural stem cells are useful for the treatment of
 CC neurological disorders, especially for repair of accidentally induced
 CC trauma in the CNS or for correction of congenital or pathological
 CC diseases of the CNS. A patient with a neurological disorder can act as a
 CC self-donor. Cells may be isolated from the patient and either sorted to
 CC extract neuroblasts, or treated in order to differentiate neuroblasts,
 CC from specific or general precursors.
 XX
 XX Sequence 2312 BP; 421 A; 739 C; 714 G; 438 T; 0 other;
 Query Match 53.2%; Score 266.2; DB 20; Length 2312;
 Best Local Similarity 77.0%; Pred. No. 2e-46;
 Matches 338; Conservative 0; Mismatches 98; Indels 3; Gaps 1;
 QY 42 GCGCGCGCGCGCGCAAACTTCGGGGGGCGCGCGCGCACTCCACCGCGCGCGC 101
 Db 259 GCGCGCGCGCGCGCGCTCTCGGGGCAAGCGCGCGCGCGCGCGCGCGCGCGC 318
 QY 102 GCGCGCGCAACCAAAACAGCCGCGACCGCGCTCAAGCGCGCCATGATGCTTATG 161
 Db 319 GCGCGCGGGGGCAAAAGGGGGGACAGACCGCTGAAGCGCCCATGAAGCGCTTATG 378
 QY 162 GTGGTCCCGCGCGCGCGCGCGAGATGAGCCAGAGAAACCCCAAGATGCACAACTGGA 221
 Db 379 GTGGTCCGGGGGCGAGCGCGCGAGATGAGCCAGAGAAATCCCAAGATGCACAACTGGA 438
 QY 222 GATAGCAAGCGCGCGCGCGAGTGAACCTTGTGGGAGAGCGAGCAACCGCGCTT 281
 Db 439 GATAGCAAGCGCGCTGGGGCGCGAGTGAAGTGTGGGCGCGAGCAACCGCGCTT 498
 QY 282 CATGACAGAGCTTAAGCGGTGCGAGCGCTGCATGAAGAGCAACCGCGATTAATA 341
 Db 499 CATGACAGAGCGCAAGCGGTGCGGGCGCTGCACATGAAGAGCAACCGCGATTAATA 558
 QY 342 CCGGCGCGCGCGGAAACCAAGAGCGCTCATGAAGAGATTAAGTACAGCTGCCGCGG 401
 Db 559 CCGGCGCGCGCGGAAAGCAAGAGCGCTCATGAAGAGCAAGTACTGCTGCCGAGG 618
 QY 402 GCTGCTGCGCGCGCGCGCAATAGCATGGGCGAGCGGGTGGCGGGCGCGCGCTGGG 461
 Db 619 GCTGCTGG--GCGCGCGCGCGCGCGCGCGCTCCCGCGTGGCGTGGGATGGG 675
 QY 462 CCGCGCGCGTGAACCAAGCGC 480
 Db 676 CCGCGCGCGTGAATCCCGCGC 694

RESULT 11
 AAX16153
 ID AAX16153 standard; CDNA to mRNA; 2376 BP.
 XX
 AC AAX16153;
 XX
 DT 22-APR-1999 (first entry)
 XX
 DE Mouse Sox1 CDNA.
 XX
 KW Mouse; Sox1; neuronal stem cell gene; neuroblastic cell; cancer;
 KW nervous system; neurological disorder; ss.
 OS Mus musculus.
 XX
 PN W09900516-A2.
 XX
 PD 07-JAN-1999.
 XX
 PF 25-JUN-1998; 98WO-GB01862.
 XX
 PR 25-JUN-1997; 97GB-0013469.
 XX
 PA (MED1-) MEDICAL RES COUNCIL.
 XX
 PI Lovell-Badge R, Pevny LH, Smith A;
 XX
 DR WPI; 1999-095759/08.
 XX
 PT Isolating neuroblastic cells from a population - by detecting the
 PT expression of the Sox1 gene in the cells and sorting the cells to
 PT isolate those cells expressing Sox1
 XX
 PS Disclosure; Page 43-44; 60pp; English.
 XX
 CC A method has been developed for isolating neuroblastic cells from a cell
 CC population. The method comprises: (a) detecting Sox1 gene expression in
 CC the cells; and (b) isolating those cells expressing Sox1. Also described
 CC is a method for producing a cell committed to the neuronal lineage,
 CC comprising: (a) transfecting a pluripotent stem cell with a genetic
 CC construct encoding Sox1 expression; (b) culturing the stem cells to
 CC differentiate into neural cells; and (c) isolating those neural cells
 CC produced. The present sequence represents mouse Sox1 cDNA. Detection
 CC of Sox1 expressing cells is important in diagnosing and treating cancers
 CC of the nervous system. Neural stem cells are useful for the treatment of
 CC neurological disorders, especially for repair of accidentally induced
 CC trauma in the CNS or for correction of congenital or pathological
 CC diseases of the CNS. A patient with a neurological disorder can act as a
 CC self-donor. Cells may be isolated from the patient and either sorted to
 CC extract neuroblasts, or treated in order to differentiate neuroblasts,
 CC from specific or general precursors.
 XX
 XX Sequence 2376 BP; 439 A; 772 C; 745 G; 420 T; 0 other;
 Query Match 52.3%; Score 261.4; DB 20; Length 2376;
 Best Local Similarity 77.6%; Pred. No. 2e-45;
 Matches 316; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
 QY 65 CCGGGGCGCGCGCGCACTCCACCGCGCGCGCGCGCGCAACCAAGAAACAGCC 124
 Db 518 CCGCGCGGGGCGCGCGGTGGGGCGCGCGCGCGCGCGCGCGCGCAACCAAGC 577
 QY 125 CGAGCGCGCTCAAGCGCGCCATGAATGCTTCAATGATGTGCTCCCGGGGAGGGCGCA 184
 Db 578 AGATGGGTGCTCAAGCGCGCCCATGAATGCTTCAATGATGTGCTCCCGGGGAGGGCGCA 637
 QY 185 AGATGGCCCAAGAGAAACCCCAAGATGCACAATCGAGATCAGCAAGCGCTGGCGCG 244
 Db 638 AGATGGCCCAAGAGAAACCCCAAGATGCACAATCGAGATCAGCAAGCGCTGGCGCG 697
 QY 245 AGTGAACCTTTGTGGAGACGAGAGCGCGCGCTTCAATCAGCAGGCTTAAGCGGCTGC 304

```
Db 698 AGTGAAGTCATGTCGAGGCCGAGACGGCCGTTTCATGACGAGGCCAAGAGACTGC 757
OY 305 GAGCGCTGCATCAGAGGACGCCCGGATTAATTAATACCGCCCGCGGAAAAACAAGA 364
Db 758 GCGCGCTGCATCAGAGGACGCCCGGATTAATTAATACCGCCCGCGGAAAAACAAGA 817
OY 365 GCGTCATGAAAGAGTAAGTACACGCTGCCCGCGCGGCTGCTGGCCCCCGCGGCAATA 424
Db 818 CGCTGCTCAAGAGGACGACTGCTGCGCGCGGCTGTAGCGCGCGCGGCGGTG 877
OY 425 GCATGCGCAGCGGGGCTCGGGGTGGCGCGGCTGGGGCGCGGGCGCTG 471
Db 878 GCGGCGCGCGCGCGCGCTGCGGCTGTGGGCGCTGGGCGGTG 924

RESULT 12
AADI1118
ID AADI1118 standard; DNA; 2509 BP.
AC AADI1118;
DT 24-SEP-2001 (first entry)
DE Human small cell lung cancer associated gene, SOX3.
XX
XX
XX Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KM leiomyosarcoma; synovial sarcoma; cytostatic; SOX3; ds.
XX
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 441..1773
FT /tag= a
FT /product= "Human SOX3 protein"
XX
XX MO200153349-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US02015.
XX
XX 21-JAN-2000; 2000US-0489101.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX WPI; 2001-457597/49.
XX P-PDB; AAE05813.
XX
XX Isolated polypeptide, used to treat or prognose a disorder
XX characterized by expression of a hCAAP e.g. cancer, is encoded by an
XX isolated nucleic acid comprising an NA Group 3 or 4 molecule -
XX
XX Claim 57; Page 98-99; 152pp; English.
XX
XX The invention relates to nucleic acids and encoded polypeptides which
XX are cancer associated antigens expressed in patients afflicted with
XX small cell lung cancer. The molecules provided by the invention can be
XX used in the diagnosis, monitoring, research or treatment of conditions
XX characterised by the expression of one or more cancer associated
XX antigens. The polypeptide is used to treat a disorder characterised by
XX expression of a hCAAP, and determine regression, progression or onset
XX of a condition characterised by expression of an abnormal amount of a
XX protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
XX are small and non-small cell lung cancer, melanoma, colon, breast, head
XX and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
XX The present sequence is a small cell lung cancer associated gene
XX designated as NY-SCILC-9, encoding human SOX3 protein.
```

```
SQ Sequence 2509 BP; 456 A; 818 C; 757 G; 478 T; 0 other:
Query Match 47.0%; Score 234.8; DB 22; Length 2509;
Best Local Similarity 73.3%; Pred. No. 6.5e-40;
Matches 318; Conservative 0; Mismatches 107; Indels 9; Gaps 1;

OY 76 GCGCGCAACTCCACCGCGCGCGCGCGCGCAACCAAAAAACGCCCGACCGCTC 135
Db 799 GCGCGCGCGCAGCAGCGGTGCTGAGCCGAGGTGCGGGGTACAGACGACCGCTGTG 858
OY 136 AAGCGCCCATGAATGCCCTTCATGGTGTGTCGCCGCGGACGCGCGCAAGATGCCCCAG 195
Db 859 AAGCGCCCATGAATGCCCTTCATGGTGTGTCGCCGCGGACGCGCGCAAAATGGCCCTG 918
OY 196 GAGAACCCCAAGATGCACACTCGGAGATCAGCAAGCGCTGGGCGCCGAGTGAACCTT 255
Db 919 GAGAACCCCAAGATGCACAACTTCTGAGATCAGCAAGCGCTGGGCGCCGAGTGAACCTG 978
OY 256 TTGTGGAGACGAGAGGAGGCGGCTTCATGAGAGGCTAAGGCGGCTGCGGCAC 315
Db 979 CTGACCGACGCGAGAGGCGACCATTCATGACGAGGCCAAGCGACTTCGCGCGTGCAC 1038
OY 316 ATGAAGAGACACCCGATTAATATACCGGCCCGCGGCGGAAACCAAGACGCTCATGAAG 375
Db 1039 ATGAAGAGATATCCGAGACTACAAAGTACCGACGCGCGGCAAGACCAAGACGCTCATGAAG 1098
OY 376 AAGGATAAGTACAGCTGCCCGCGGCGTGTGCGCCCGCGGCAATACATGCGCAGC 435
Db 1099 AAGGATAAGTACTCTCCTGCCACGCGCTCTCTCTCGGCTCGCGCGCGCGCGCGCC 1158
OY 436 GGGGTGCGG-----GTGGGCGCGGCGCTGGGCGGCGGTGAACCAAGCGCATGAC 486
Db 1159 GCTGCCGCGCGCGCGACCGCTGCCCGCAGCAGTCGCTGGGCGTGGGCGCGCGCTGGAC 1218
OY 487 AGTTACGCGCACAT 500
Db 1219 ACGTACAGCACAGCT 1232

RESULT 13
AADI1119
ID AADI1119 standard; DNA; 8372 BP.
XX
XX AADI1119;
XX
XX 24-SEP-2001 (first entry)
XX
XX Human small cell lung cancer associated gene, SOX21.
XX
XX Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KM leiomyosarcoma; synovial sarcoma; cytostatic; SOX21; ds.
XX
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1167..1997
FT /tag= a
FT /product= "Human SOX21 protein"
XX
XX MO200153349-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US02015.
XX
XX 21-JAN-2000; 2000US-0489101.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
```

XX WPI: 2001-457597/49.
DR p-PsDB; AAE05814.
XX
PT Isolated polypeptide, used to treat or prognose a disorder
PT characterized by expression of a hCAP e.g. cancer, is encoded by an
PT isolated nucleic acid comprising an NA Group 3 or 4 molecule -
XX
PS Claim 57; Page 100-105; 152pp; English.

CC The invention relates to nucleic acids and encoded polypeptides which
CC are cancer associated antigens expressed in patients afflicted with
CC small cell lung cancer. The molecules provided by the invention can be
CC used in the diagnosis, monitoring, research or treatment of conditions
CC characterised by the expression of one or more cancer associated
CC antigens. The polypeptide is used to treat a disorder characterised by
CC expression of a hCAP, and determine regression, progression or onset
CC of a condition characterised by expression of an abnormal amount of a
CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
CC are small and non-small cell lung cancer, melanoma, colon, breast, head
CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma.
CC The present sequence is a small cell lung cancer associated gene
CC designated as NY-SCIC-10, encoding human SOX21 protein.
XX

SQ Sequence 8372 BP; 1996 A; 2143 C; 2104 G; 2124 T; 5 other;

Query Match 45.0%; Score 225; DB 22; Length 8372;
Best Local Similarity 71.6%; Pred. No. 7.5e-38;
Matches 310; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

QY 67 GGGGGGGGGGGGGGCACTCACCAGGGGGGGGGGGGGGCAACCAAGAAAGAGCCG 126
DB 1122 GAGCGGG 1181
QY 127 GAGCGGG 186
DB 1182 GAGCGGG 1241
QY 187 ATGCGCCAG 246
DB 1242 ATGCGCCAG 1301
QY 247 TGGAACTTTTGGGAG 306
DB 1302 TGGAACTTTTGGGAG 1361
QY 307 GCGCTGCACATGAAG 366
DB 1362 GCGCTGCACATGAAG 1421
QY 367 CMCATGAAG 423
DB 1422 CMCATGAAG 1481
QY 424 AGCATGGCAG 483
DB 1482 AGCATGGCAG 1541
QY 484 GACAGTTAGCGCG 496
DB 1542 GTGCTGAGTGC 1554

RESULT 14
AAS87738
ID AAS87738 standard; cDNA; 2378 BP.
XX AAS87738;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23542.
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YF;
XX
DR WPI: 2001-639362/73.
XX
DR p-PsDB; ABG23551.

Claim 1; SEQ ID NO 23542; 103pp; English.

CC The invention relates to isolated polynucleotide (i) and
CC polypeptide (ii) sequences. (i) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (ii). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (i) is useful in gene therapy techniques
CC to restore normal activity of (ii) or to treat disease states involving
CC (ii). (ii) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (ii) and its binding partners are useful in medical
CC imaging of sites expressing (ii). (i) and (ii) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC atftp.wipo.int/pub/published_pcl_sequences.
XX

SQ Sequence 2378 BP; 590 A; 708 C; 664 G; 416 T; 0 other;

Query Match 44.0%; Score 220; DB 23; Length 2378;
Best Local Similarity 69.0%; Pred. No. 7.5e-37;
Matches 334; Conservative 0; Mismatches 140; Indels 10; Gaps 2;

QY 27 GGAGCGAGCTGAAG 86
DB 733 GGAGCGAGCGAGCAAGATGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792
QY 87 CACCGGG 145
DB 793 AGCGGTGTGGAG 852
QY 146 TGAATGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 205
DB 853 TGAAGGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
QY 206 AGATGCACACTCGGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265
DB 913 AGATGCACAAATTCGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972
QY 266 CGGAGAGAGGG 325


```

Db 973 CCGAAGCGACCATTCATCGACGAGCCAGCACTTGGCCGCTGCACATGAAGAGT 1032
QY 326 ACCGGATTATTAATATACGGGCCCCCGGAAACCAAGACCTCATGAAGAAGATTAAGT 385
Db 1033 ATCCGGACTTACAAAGTACGACCGCGCCGCAAGACCAAGACGCTGCTCAAGAAAGATTAAGT 1092
QY 386 ACACGCTGCCCGCGGGGCTGTGGCCCGCGGCAATAGCATATGGGAGCGGGGTGGG- 444
Db 1093 ACTCCCTGCCACGCGGCTCTCTGCTCCCGGTGCGCGGCCCCCGCGCTGCGCGG 1152
QY 445 -----GTGGGCGCGCGGCTGTGGCGGCGGTGAACCGCATGACATTAACGCGC 496
Db 1153 CCGCAGCGCGTGCAGCGCAGAGTCCGTGGGCGGTGGCGCGCTGACACGTACACGC 1212
QY 497 ACAT 500
Db 1213 ACGT 1216

```

RESULT 15

AAS61625
ID AAS61625 standard; cDNA: 412 BP.

XX AC AAS61625;

XX DT 29-JAN-2002 (first entry)

DE Lung small cell carcinoma antigen, cDNA #166.

XX Human; cytosolic; antitumour; lung small cell cancer antigen;

KW tumour; lung cancer; ss.

XX OS Homo sapiens.

PN W0200177168-A2.

XX PD 18-OCT-2001.

XX PF 11-APR-2001; 2001WO-US11859.

XX PR 11-APR-2000; 2000US-196780P.

XX PR 21-JUN-2000; 2000US-213361P.

XX PR 01-SEP-2000; 2000US-229763P.

XX PR 05-SEP-2000; 2000US-230629P.

XX PR 14-SEP-2000; 2000US-232655P.

XX PR 19-DEC-2000; 2000US-257037P.

XX PR 08-JAN-2001; 2001US-260796P.

XX PA (CORI-) CORIXA CORP.

XX PI Lodes MJ, Wang T, Mohamath R, Indirias CY;

XX DR WPI; 2002-010896/01.

XX PT Lung tumour polynucleotide and polypeptides useful in therapy and

XX PT diagnosis of cancer especially lung cancer -

XX PS Claim 1; Page 180; 295pp; English.

XX CC The invention relates to novel isolated lung small cell cancer antigen
 CC polynucleotides (I) and polypeptides (II) used in a method of detecting
 CC cancer in a patient. The method is optionally performed by
 CC utilising oligonucleotides (III), where the biological sample
 CC from the patient is contacted with (III), detecting the amount of
 CC polynucleotide hybridised to (III) in the sample and comparing the
 CC amount of polynucleotide to a predetermined cut-off value and thereby
 CC determining cancer in a patient. (I), (II) or antigen-presenting cells
 CC expressing (II) is useful for stimulating and/or expanding T cells
 CC specific for a tumour protein. The method comprises contacting T cells
 CC with one of the components under conditions to permit the stimulation
 CC and/or expansion of the cells. A composition comprising (I) is useful for
 CC stimulating an immune response in a patient and for inhibiting the

CC development of a cancer especially lung cancer in a patient. An
 CC isolated T cell population is useful for removing tumour cells from the
 CC biological sample and for inhibiting the development of cancer in a
 CC patient. AAS61460-AAS61874 represent novel human lung small cell
 CC cancer antigen coding sequences of the invention.

SQ Sequence 412 BP; 83 A; 133 C; 144 G; 51 T; 1 other;

Query Match 43.6%; Score 217.8; DB 24; Length 412;

Best Local Similarity 75.7%; Pred. No. 1.9e-36;

Matches 283; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

```

QY 126 GGACCGGCTCAAGGGGCCCATGAATGCTTCATGGTGTGGTCCGCGGCGACGGCGCA 185
Db 15 GGACCACTCAAGGGGCCCATGAATGCTTCATGGTGTGGTCCGCGGCGACGGCGCA 74
QY 186 GATGGCCAGAGAAACCCCAAGATGACACACTCGAGATCAGACCAAGCGCTGGGCGCA 245
Db 75 GATGGCCAGAGAAACCCCAAGATGACACACTCGAGATCAGACCAAGCGCTGGGCGCA 134
QY 246 GTGGAACCTTTGTCGGAGAGGAGAACGCGCGCTTCATCCAGAGCTAACGCGCTGCG 305
Db 135 GTGGAACCTGCTCACAGAGTGGAGAACGCGCGCTTCATCCAGAGGCGCCAGCTTACG 194
QY 306 AGCGCTGCATGAAGAGCAACCGGATTTAATATACCGGCGCGGCGGAAACCAAGAC 365
Db 195 CGCATGACATGAAGAGCAACCGGATTTAATATACCGGCGCGGCGGCGGCGGCA 254
QY 366 GCTCATGAAGAGATTAATATACAGCTGCCG--GCGGGCTGCTGGCGCGCGCGGCGCA 422
Db 255 GCTCATGAAGAGATTAATATACAGCTGCCGCTTCCGCTTACAGGCGCTGGGCGGCGGCA 314
QY 423 TAGCATGGCGAGCGGGGCTGGGCGCGCGCTGGGCGCGGCGGCGGCGGCGGCGAT 482
Db 315 CGCGGACACCTTGGCTCAAGGCGGCGCGGCTGACCGGCGGCGGCGGCGGCGGCT 374
QY 483 GGACATTAAGCGCGC 496
Db 375 GGTGCTGAGTGC 388

```

Search completed: October 10, 2002, 14:28:33
 Job time : 92 secs

